REMARKS

This paper is being filed in response to the Office Action dated June 3, 2003 that was issued in the above-identified application. Applicants respectfully request continued examination of the instant application pursuant to 35 U.S.C. § 132(a) and 37 C.F.R. § 1.114(a)(2) and enclose herewith the fee required pursuant to 37 C.F.R. § 1.17(e). Applicants also request a three-month extension of time and enclose the fee required under 37 C.F.R. §1.17(a)(3). Applicants further enclose herewith a Third Substitute Sequence Listing in paper and computer readable form in accordance with 37 C.F.R. §§ 1.821 to 1.825 and a Terminal Disclaimer under 37 C.F.R. § 1.321(b). Applicants respectfully request reconsideration of the above-identified application in light of the amendments and remarks presented in the instant Amendment.

Claims 42-51, 53, 55-56, 82, and 85-86 are pending. Claims 42-45, 53, and 55 have been amended. Dependent claim 42 has been amended to correspond to independent claim 43 and to recite "isolated or purified". Independent claim 43 has been amended to correspond to dependent claim 42. Claims 44-45 and 55 have been amended to refer to claim 42 instead of claim 43. Therefore, these amendments do not constitute new matter. Amended claim 53 is supported by the specification as originally filed, for example, by Example 10 and, therefore, does not constitute new matter. Upon entry of the instant Amendment, claims 42-51, 53, 55-56, 82, and 85-86 will continue to be pending.

As a preliminary matter, Applicants thank the Examiner for withdrawing earlier objections to the specification and claims. Applicants further thank the Examiner for withdrawing many of the earlier rejections to the claims including rejections that the claimed invention lacks utility, is not enabled by the specification, is indefinite, and is anticipated by Campbell (1993, *J. Clin. Microbiol.* 31:2255-2262), Smith (1998, *Toxicol.* 36:1539-1548),

Halpern (1993, JBC 268:11186-11192), Whelan (Accession No. M81186), Jung (1992, *FEMS Microbiol. Lett.* 91:69-72), and Williams (U.S. Patent No. 5,919,665).

Amendments Are Fully Supported

The specification has been amended to recite Accession Nos. X52066 and M81186. This amendment is supported by the specification as originally filed. For example, the specification originally cited Thompson et al., 1990, *European Journal of Biochemistry* 189:73-81 at page 12, lines 9-10. The botulinum neurotoxin serotype A sequence disclosed in Figure 3 (pp. 76-77) of this document was deposited in Genbank and assigned Accession No. X52066. *See* Accession No. X52066 (annotations recite Thompson et al. as the only "reference")(Exhibit 1). Additionally, the specification originally cited Whelan et al., 1992, *Applied and Environmental Microbiology* 58:2345-2354 at page 13, line 1. This article recites Accession No. M81186 at page 2346, second column, second full paragraph.

SEQ ID NOS: 7, 37, 39, 40, 41, and 42 have been amended herein. SEQ ID NO:7 has been amended to agree with Figure 4 as originally filed with the instant application. Therefore, this amendment does not constitute new matter.

SEQ ID NO:37 has been amended to correspond to the sequence shown in Figure 2 of U.S. Patent Application No. 08/123,975 by Middlebrook et al. filed on September 21, 1993 (hereinafter "the '975 application") and to which the instant application claims priority. Therefore, this amendment does not constitute new matter.

SEQ ID NO:39 has been amended to correspond to the sequence shown in Figure 4 of the '975 application. Therefore, this amendment does not constitute new matter.

SEQ ID NO:40 has been amended to agree with Figure 3 on page 2349 of Whelan (Accession No. M81186)(hereinafter "Whelan"). SEQ ID NO:41 has been amended to agree

with Figure 3 on pages 76-77 of Thompson (Accession No. X52066)(hereinafter "Thompson"). SEQ ID NO:42 has been amended to agree with Whelan. Applicants assert that Thompson and Whelan were available to those of ordinary skill in the art at the time the instant application was filed. Specifically, the National Center for Biotechnology Information indicates that Thompson was "first seen" on April 21, 1993 (Exhibit 1) and Whelan was "first seen" on April 26, 1993 (Exhibit 2). In addition, the specification has been amended to specifically incorporate Thompson and Whelan by reference. In this regard, Applicants respectfully invite the Examiner's attention to page 44, line 18-20 of the specification, which states "All publications and patent applications are herein incorporated by reference to the same extent as if each individual publication or patent application was specifically and individually indicated to be incorporated by reference.." Therefore, these amendments do not constitute new matter.

Declarations

Applicants submit herewith a Third Substitute Sequence Listing in paper and computer readable form. I hereby state that the content of the paper and computer readable copies of the Third Substitute Sequence Listing submitted in accordance with 37 C.F.R. §1.821(c) and (e), are the same. I hereby state that the content of the paper and computer readable copies of the Third Substitute Sequence Listing, submitted in accordance with 37 C.F.R. §1.821(g), herein does not include new matter.

Applicants Third Substitute Sequence Listing corrects typographical errors in the sequences presented in the original application and in the Amendment and Substitute Sequence Listing filed on March 5, 2002. Applicants enclose herewith six sequence alignments for the Examiner's review. The original sequence, the sequence substituted on March 5, 2002 (hereinafter "the Substitute sequence"), and the sequence as amended herein (hereinafter "the

Amended sequence") are aligned along with the sequence relied upon to support the instant amendment. Positions where the sequences differ are highlighted, while positions where all aligned sequences match are marked with an asterix. These alignments have been prepared with ClustalW 1.74 accessed at <http://www.ch.embnet.org/software/ClustalW.html > with default settings.

Applicants have amended the specification to incorporate the Thompson and Whelan sequences in the sequence listing respectively as SEQ ID NOS:43 and 44. Both Thompson and Whelan sequences were incorporated by reference into the instant application. *See* page 44, lines 18-20. I hereby declare that SEQ ID NO: 40, as amended herein, is the same as Whelan amino acids 853 to 1291. I hereby declare that SEQ ID NO: 41, as amended herein, is the same as Thompson amino acids 449 to 1296. I hereby declare that SEQ ID NO: 42, as amended herein, is the same as Whelan amino acids 442 to 1291.

Claims Are Arranged in Proper Sequence

Claim 42 has been objected to as allegedly depending on a later claim. Applicants have amended claim 42 to replace claim 43 and vice-versa. Therefore, upon entry of the instant amendment, claims 42 and 43 will be in proper sequence. Applicants, therefore, respectfully request withdrawal of this objection.

Claims Relate to Patentable Subject Matter

Claims 39-43 (*sic*, 42-43), 45-47, and 55-56 have been rejected under 35 U.S.C. § 101 as allegedly directed to non-statutory subject matter. The Examiner has alleged that natural variation in nucleotide and amino acid levels for the same or equivalent proteins is common in nature.

Applicants traverse this rejection and assert that the claims, as amended herein, are drawn to patentable subject matter. Applicants, therefore, respectfully request withdrawal of this rejection.

Claims Are Supported By Adequate Written Description

Claims 42-51, 53-56 (*sic*, 53, 55-56), 82, and 85-86 have been rejected under 35 U.S.C. § 112, first paragraph as allegedly lacking sufficient written description. The Examiner has alleged that Example 8 and Figure 4 do not define a representative number of species of the instantly claimed genus of nucleic acid molecules.

Applicants traverse this rejection and assert that the claims, as amended herein, are supported by an adequate written description. The Examiner has alleged that the phrase "a nucleic acid comprising a nucleic acid sequence" encompasses variant nucleic acid sequences. This phrase has been omitted. Applicants, therefore, respectfully request withdrawal of this rejection.

Claims Are Clear and Definite

Claim 43 has been rejected under 35 U.S.C. § 112, second paragraph as allegedly indefinite in reciting "said amino acid sequence comprising at least one immunogenic epitope". The Examiner has alleged that it is unclear whether the immunogenic epitope is inherent to SEQ ID NO:8 or heterologous to SEQ ID NO:8 since the term "having" is construed to be open.

Applicants traverse this rejection and assert that it is clear that the immunogenic epitope is part of SEQ ID NO:8. Claim 42 has been amended to replace claim 43 and vice-versa. Applicants respectfully submit that the antecedent basis for the phrase "said amino acid sequence comprising at least one immunogenic epitope" recited in claim 42 is the immediately prior

phrase, namely, "the amino acid sequence **of** SEQ ID NO:8" (emphasis added). Thus, it is clear that the immunogenic epitope is part of SEQ ID NO:8. Accordingly, Applicants respectfully request withdrawal of this rejection.

Claim 53 has been rejected under 35 U.S.C. § 112, second paragraph as allegedly indefinite. The Examiner has alleged that the method steps are no commensurate in scope with the preamble.

Applicants traverse this rejection and assert that claim 53, as amended herein, is clear and definite. Both the preamble and the method step recite "isolating." Applicants, therefore, respectfully request withdrawal of this rejection.

Specification and Claims Are Free of New Matter

Applicant's Preliminary Amendment and Substitute Sequence Listing mailed on March 5, 2003 has been objected to under 35 U.S.C. § 132 as allegedly introducing new matter.

1. SEQ ID NO:37

The Examiner has alleged that Figures 1, 2, and 3 do not support the amendments to SEQ ID NO:37. Without acquiescing in this rejection, Applicants have amended SEQ ID NO:37 to correct errors. This amendment is supported by Figure 2 of the '975 application to which the instant application claims priority.

2. SEQ ID NOS: 7 and 39

The Examiner has alleged that Figure 4 (SEQ ID NO:7) do not support the amendments to SEQ ID NO:39 and vice versa. Without acquiescing in this rejection, Applicants have amended SEQ ID NOS: 7 and 39 to correct errors. These amendments are respectively

supported by the original Figure 4 submitted with the instant application and Figure 4 of the '975 application.

3. SEQ ID NO:40

The Examiner has alleged that Accession No. M81186 and SEQ ID NO:8 do not support the amendments to SEQ ID NO:40. Without acquiescing in this rejection, Applicants have amended SEQ ID NO:40 to correct errors. This amendment is supported by Figure 3 of Whelan (Accession No. M81186), which was incorporated by reference into the instant application.

4. SEQ ID NO:41

The Examiner has alleged that Figures 1, 2, and 3 and SEQ ID NO:38 do not support the amendments to SEQ ID NO:41. Without acquiescing in this rejection, Applicants have amended SEQ ID NO:41 to correct errors. This amendment is supported by Figure 3 of Thompson (Accession No. X52066), which was incorporated by reference into the instant application.

5. SEQ ID NO:42

The Examiner has alleged that Figure 4 (SEQ ID NO:8) and Accession No. M81186 do not support the amendments to SEQ ID NO:42. Without acquiescing in this rejection, Applicants have amended SEQ ID NO:42 to correct errors. This amendment is supported by Figure 3 of Whelan (Accession No. M81186), which was incorporated by reference into the instant application.

Thus, all of the amendments to the sequence listing are fully supported and, therefore do not constitute new matter. Applicants, therefore, respectfully request withdrawal of this rejection.

Claims Do Not Represent Double Patenting

Claims 45, 48, 53, and 82 have been rejected under the judicially-created doctrine of obviousness-type double patenting as allegedly obvious over U.S. Patent No. 6,495,143 issued to Lee JS et al. on December 17, 2002. The Examiner indicated that a timely filed terminal disclaimer would overcome this rejection.

Applicants respectfully invite the Examiner's attention to the Terminal Disclaimer enclosed herewith. A copy of the executed Assignment is also enclosed herewith.

In summary, Applicants believe that all pending claims are in condition for allowance and respectfully solicit prompt favorable action.

In summary, Applicants believe that all pending claims are in condition for allowance and respectfully solicit prompt favorable action.

Applicants enclose herewith the fee required under 37 C.F.R. §1.17(e) and §1.17(a)(3). Although Applicants do not believe that any additional fees are required with this paper, the Commissioner is hereby authorized to charge any fees occasioned by this submission not otherwise enclosed herewith to Deposit Account No. 02-4377. Please credit any overpayment of fees associated with this filing to the above-identified deposit account. A duplicate of this page is enclosed.

Respectfully submitted,

BAKER BOTTS, L.L.P.

December 3, 2003

Rochelle K. Seide
PTO Reg. No. 32,300

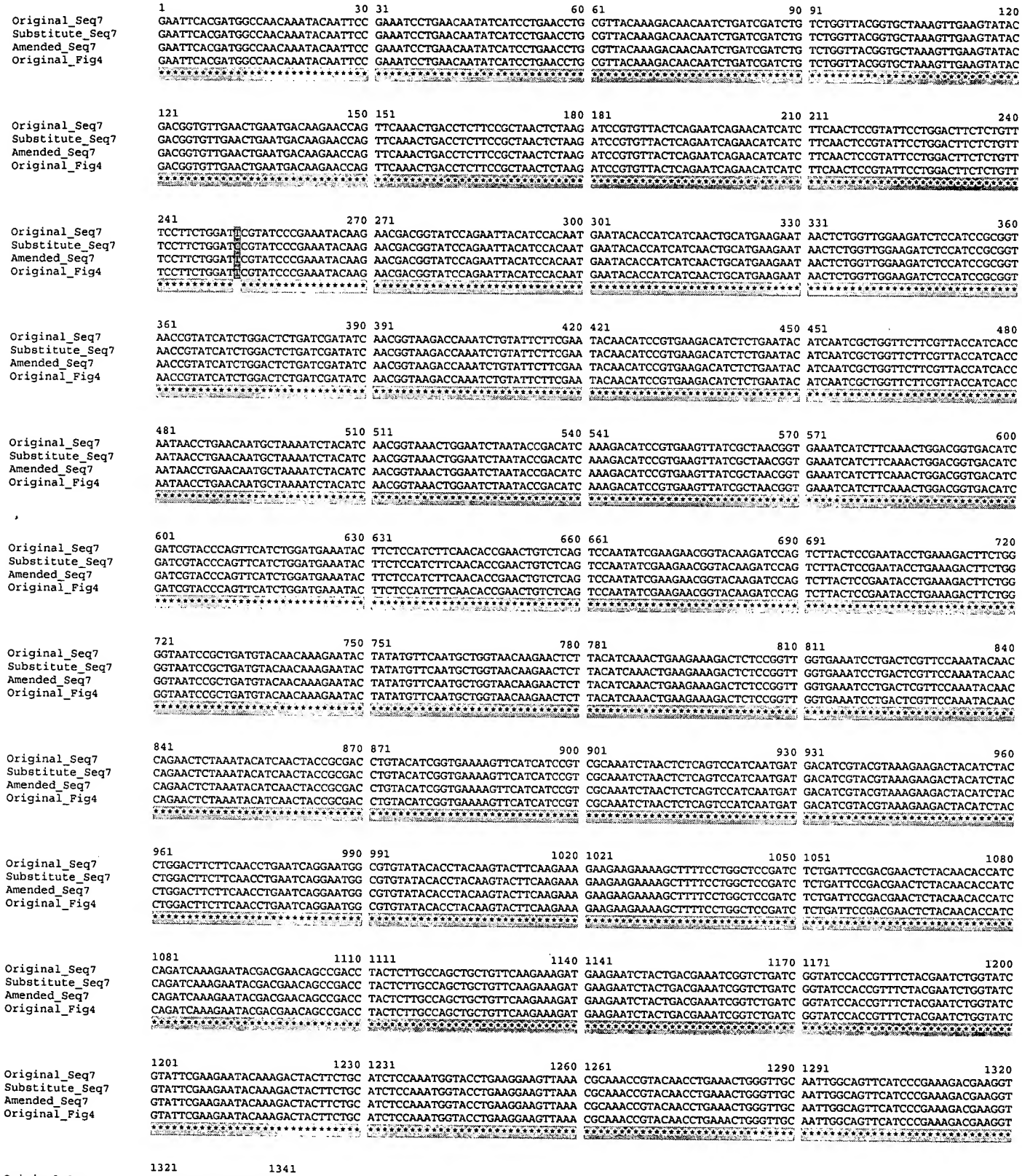
Carmella L. Stephens
PTO Reg. No. 41,328
Attorneys for Applicants

Guy F. Birkenmeier
PTO Reg. No. 52,622
Agent for Applicants

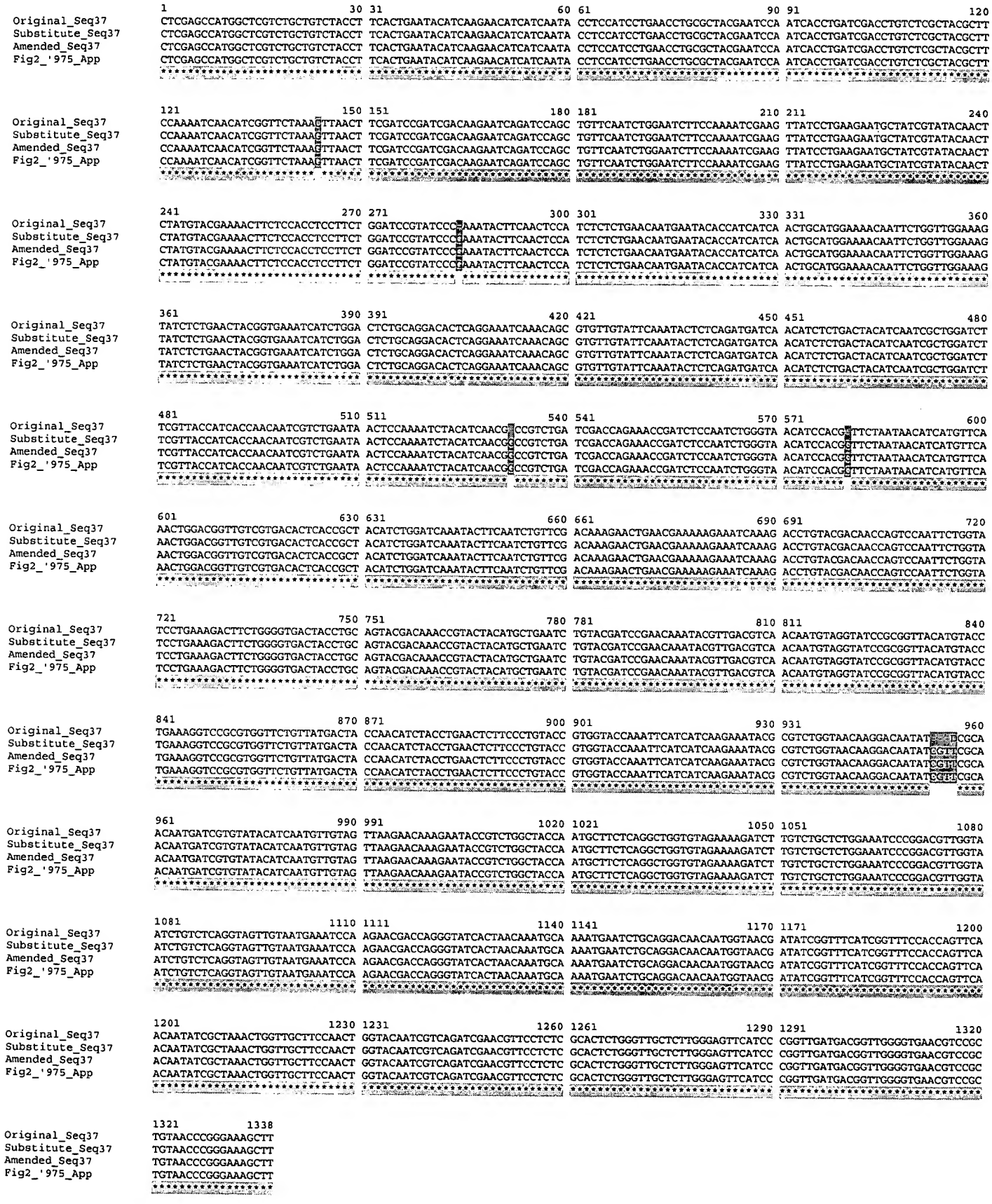
BAKER BOTTS, L.L.P.
30 Rockefeller Plaza
New York, NY 10112
(212) 408-2500

Enclosures

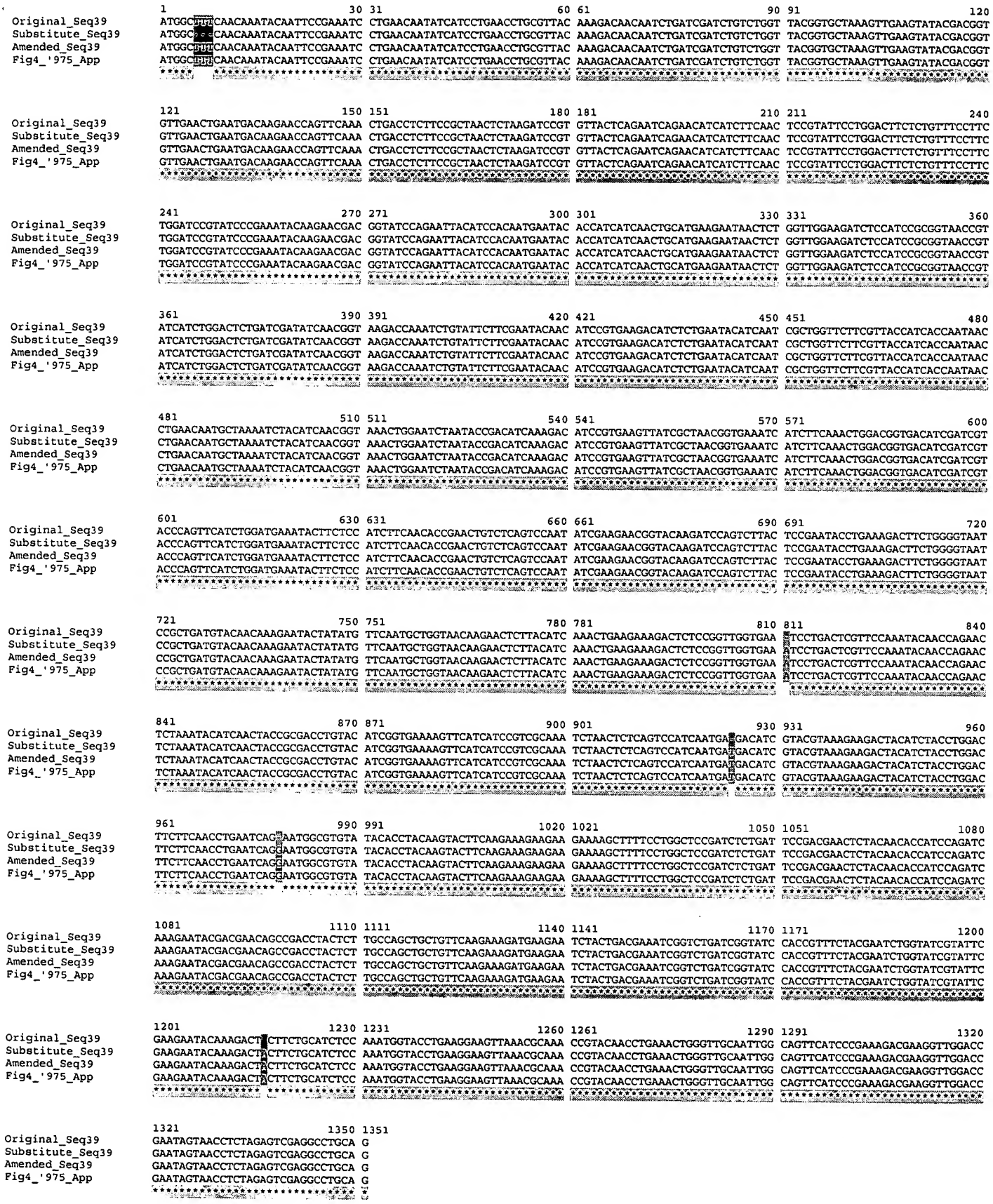
ALIGNMENT 1: SEQ ID NO:7



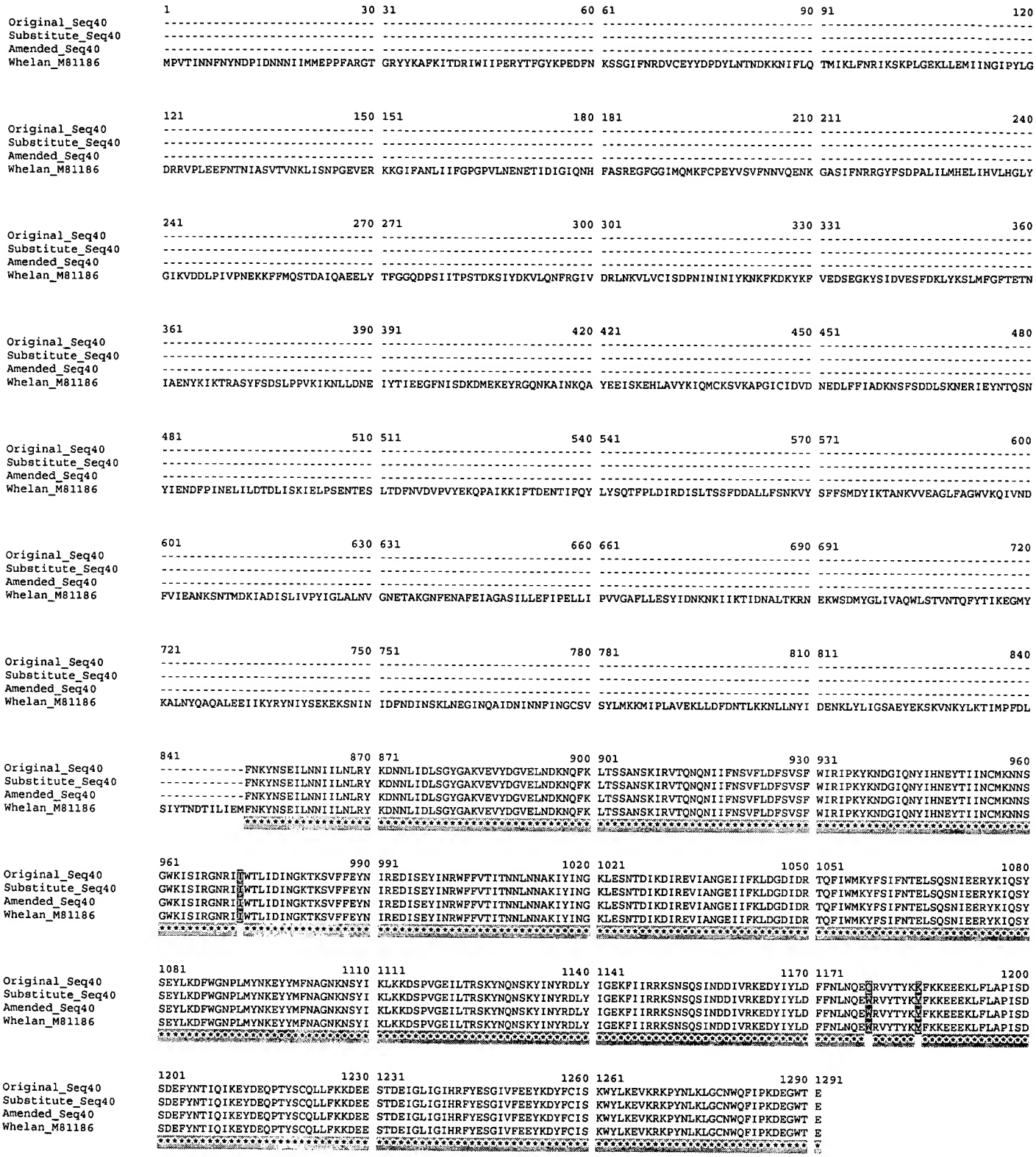
ALIGNMENT 2: SEQ ID NO:37



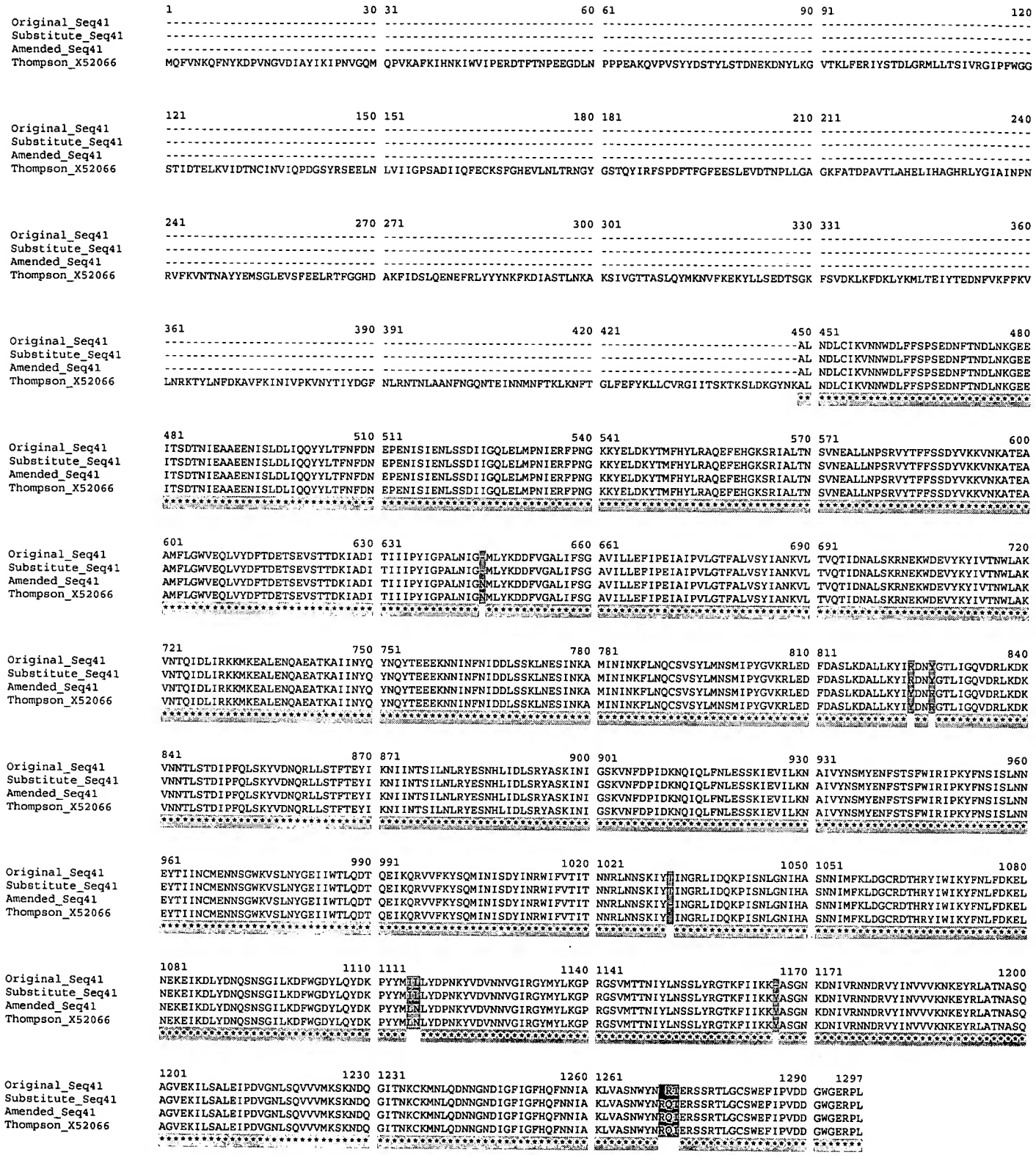
ALIGNMENT 3: SEQ ID NO:39



ALIGNMENT 4: SEQ ID NO:40



ALIGNMENT 5: SEQ ID NO:41



ALIGNMENT 6: SEQ ID NO:42

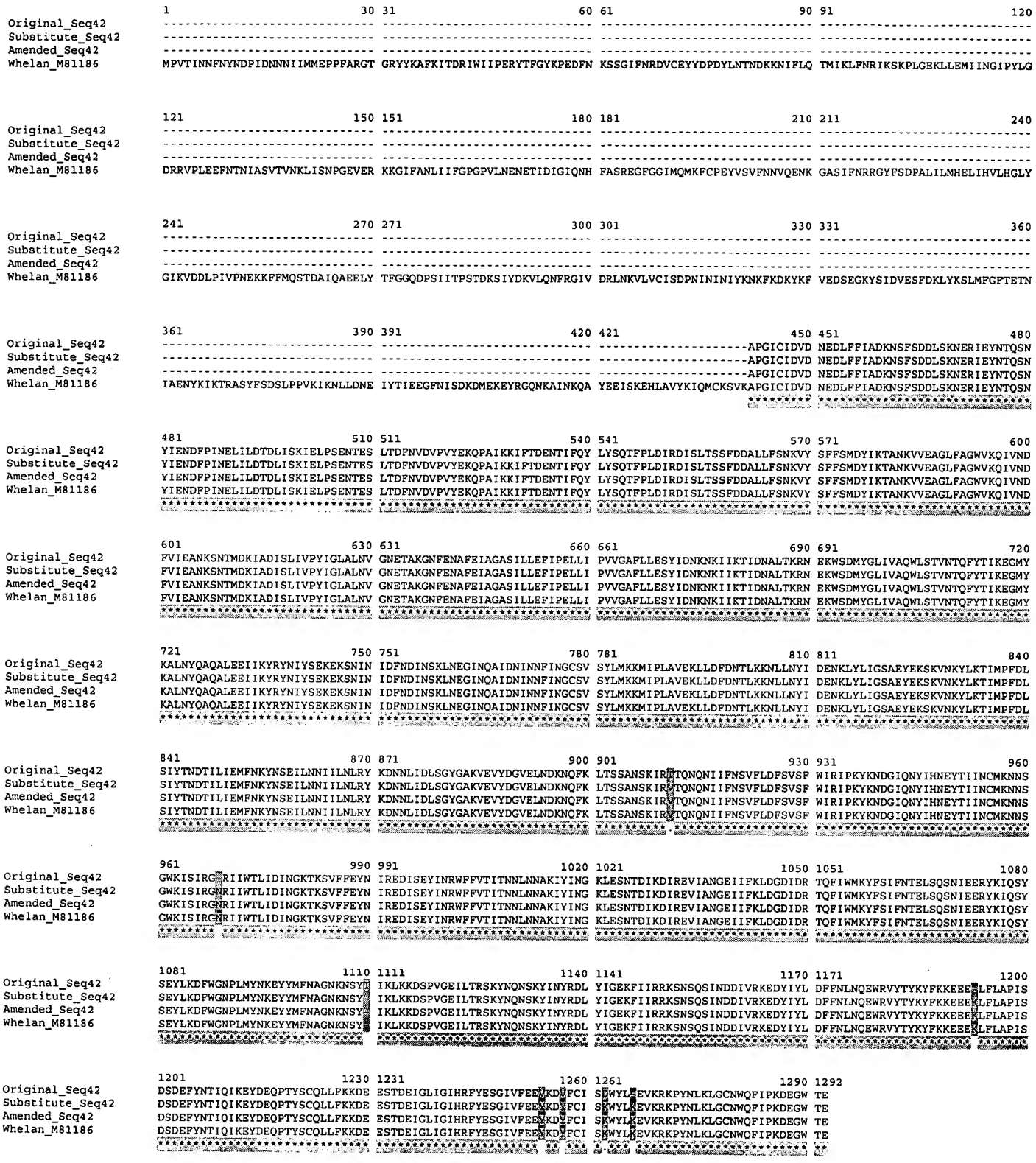


EXHIBIT 1

NCBI

Sequence Revision History

Show difference between I and II as GenBank/GenPept

Entrez

Search for Genes
LocusLink provides curated information for human, fruit fly, mouse, rat, and zebrafish

Help|FAQ

Batch Entrez: Upload a file of GI or accession numbers to retrieve protein or nucleotide sequences

Check sequence revision history

How to create WWW links to Entrez

LinkOut

Cubby

Related resources

BLAST

Reference sequence project

LocusLink

Clusters of orthologous groups

Protein reviews on the web

Revision history for X52066

GI	Version	Update Date	Status	I	II
40381	1	Aug 8 2003 11:51	Live		
40381	1	Oct 17 2002 3:35	Dead		
40381	1	Mar 8 1999 4:30	Dead		
40381	1	May 26 1996 2:31	Dead		
40381	1	May 21 1995 4:23	Dead		
40381	1	Nov 30 1994 11:08	Dead		
40381	1	Aug 31 1993 1:05	Dead		
40381	1	Apr 21 1993 8:19	Dead		

Accession X52066 was first seen at NCBI on Apr 21 1993 8:19

Search Nucleotide for Go Clear

Limits Preview/Index History Clipboard Details

Display default Show: 20 Send to File Get Subsequence Features

1: X52066. Clostridium botul...[gi:40381]

Links

```
LOCUS       CBBOTAG                 4292 bp    DNA     linear   BCT 12-SEP-1993
DEFINITION  Clostridium botulinum botA gene for type A neurotoxin.
ACCESSION   X52066 X52088
VERSION     X52066.1  GI:40381
KEYWORDS    botA gene; neurotoxin; secreted protein.
SOURCE      Clostridium botulinum
  ORGANISM  Clostridium botulinum
            Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
            Clostridium.
REFERENCE   1  (bases 1 to 4292)
  AUTHORS   Thompson,D.E., Brehm,J.K., Oultram,J.D., Swinfield,T.J.,
            Shone,C.C., Atkinson,T., Melling,J. and Minton,N.P.
  TITLE     The complete amino acid sequence of the Clostridium botulinum type
            A neurotoxin, deduced by nucleotide sequence analysis of the
            encoding gene
  JOURNAL   Eur. J. Biochem. 189 (1), 73-81 (1990)
  MEDLINE   90235864
   PUBMED   2185020
REFERENCE   2  (bases 1 to 4292)
  AUTHORS   Minton,N.P.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-JAN-1990) Minton N.P., PHLS Centre for Applied
            Microbiology & Research, Molecular Genetics Group, Division of
            Biotechnology, Porton Down, Salisbury SP4 0JG Wiltshire, U K
FEATURES             Location/Qualifiers
     source          1..4292
                     /organism="Clostridium botulinum"
                     /mol_type="genomic DNA"
                     /strain="sub sp. type A, NCTC2916"
                     /db_xref="taxon:1491"
     misc_feature    16..64
                     /note="dyad symmetry"
     promoter        19..24
                     /note="-35 region"
     promoter        41..46
                     /note="-10 region"
     RBS             63..68
                     /note="ribosome binding site"
     CDS             77..3967
                     /note="unnamed protein product; type A neurotoxin (AA
                     1-1296)"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA36289.1"
                     /db_xref="GI:40382"
                     /db_xref="GOA:P10845"
                     /db_xref="SWISS-PROT:P10845"
                     /translation="MQFVNKQFNYKDPVNGVDIAYIKIPNVGQMQPVKAFKIHNKIWV
```

IPERDTFTNPEEGDLNPPPEAKQVPVSYYDSTYLSTDNEKDNYLKGVTKLFERIYSTD
LGRMLLTSIVRGIPFWGGSTIDTELKVIDTNCINVIQPDGSYRSEELNLVIIGPSADI
IQFECKSFGHEVLNLTRNGYGSTQYIRFSPDFTFGFEESLEVDTNPLLGAGKFATDPA
VTLAHELIHAGHRLYGIAINPNRVFKVNTNAYYEMSGLEVSFEELRTFGGHDAKFIDS
LQENEFRLYYYNKFKDIASTLNKAKSIVGTTASLQYMKNVFKEKYLLSEDTSGKFSVD
KLKFDKLYKMLTEIYTEDNFVKFFKVLNRKTYLNFDKAVFKINIVPKVNYTIYDGFNL
RNTNLAANFNGQNTEINNMNFTKLKNFTGLFEFYKLLCVRGIITSKTKSLDKGYNKAL
NDLCIKVNNWDLFFSPSEDNFTNDLNKGEEITSDTNIEAAEENISLDLIQQYYLTFNF
DNEPENISIENLSSDIIGQLELMPNIERFPNGKKYELDKYTMFHYLRAQEFEHGKSRI
ALTNSVNEALLNPSRVYTFFSSDYVKKVNKATEAAMFLGWVEQLVYDFTDETSEVSTT
DKIADITIIIPYIGPALNIGNMLYKDDFVGALIFSGAVILLEFIPEIAIPVLGTFALV
SYIANKVLTVQTIDNALSKRNEKWDEVYKYIVTNWLAKVNTQIDLIRKKMKEALENQA
EATKAIINYQYNQYTEEEKNNINFNIDDLSSKLNESINKAMININKFLNQCSVSYLMN
SMIPYGVKRLEDFDASLKDALLKYIYDNRGTLIGQVDRLKDKVNNTLSTDIPFQLSKY
VDNQRLLSTFTEYIKNIINTSILNLRYESNHLIDLSRYASKINIGSKVNFDPIDKNQI
QLFNLESSKIEVILKNAIVYNSMYENFSTSFWIRIPKYFNSISLNNEYTIINCMENNS
GWKVSLNYGEIIWTLQDTQEIKQRVVFKYSQMINISDYINRWIFVTITNNRLNNSKIY
INGRLIDQKPISNLGNIHASNNIMFKLDGCRDTHRYIWIKYFNLFDKELNEKEIKDLY
DNQSNSGILKDFWGDYLQYDKPYYMLNLYDPNKYVDVNNVGIRGYMYLKGPRGSVMTT
NIYLNSSLYRGTKFIIKKYASGNKDNIVRNNDRVYINVVVKNKEYRLATNASQAGVEK
ILSALEIPDVGNLSQVVVMKSKNDQGITNKCKMNLQDNNGNDIGFIGFHQFNNIAKLV
ASNWYNRQIERSSRTLGCSWEFIPVDDGWGERPL"

misc_feature 4042..4087
/note="dyad symmetry"

misc_feature 4104..4136
/note="dyad symmetry"

misc_feature 4158..4188
/note="dyad symmetry"

ORIGIN

```
   1 tcaaagtatt tgtatttatg gtcatttaaa taattaataa tttaattaat tttaaatatt
  61 ataagaggtg ttaaatatgc aatttgttaa taaacaattt aattataaag atcctgtaaa
 121 tggtgttgat attgcttata taaaaattcc aaatgtagga caaatgcaac cagtaaaagc
 181 ttttaaaatt cataataaaa tatgggttat tccagaaaga gatacattta caaatcctga
 241 agaaggagat ttaaatccac caccagaagc aaaacaagtt ccagtttcat attatgattc
 301 aacatattta agtacagata atgaaaaaga taattattta aagggagtta caaaattatt
 361 tgagagaatt tattcaactg atcttggaag aatgttgtta acatcaatag taaggggaat
 421 accatttttgg ggtggaagta caatagatac agaattaaaa gttattgata ctaattgtat
 481 taatgtgata caaccagatg gtagttatag atcagaagaa cttaatctag taataatagg
 541 accctcagct gatattatac agtttgaatg taaaagcttt ggacatgaag ttttgaatct
 601 tacgcgaaat ggttatggct ctactcaata cattagattt agcccagatt ttacatttgg
 661 ttttgaggag tcacttgaag ttgatacaaa tcctctttta ggtgcaggca aatttgctac
 721 agatccagca gtaacattag cacatgaact tatacatgct ggacatagat tatatggaat
 781 agcaattaat ccaaataggg tttttaaagt aaatactaat gcctattatg aaatgagtgg
 841 gttagaagta agctttgagg aacttagaac atttgggggа catgatgcaa agtttataga
 901 tagtttacag gaaaacgaat ttcgtctata ttattataat aagtttaaag atatagcaag
 961 tacacttaat aaagctaaat caatagtagg tactactgct tcattacagt atatgaaaaa
1021 tgtttttaaa gagaaatatc tcctatctga agatacatct ggaaaatttt cggtagataa
1081 attaaaattt gataagttat acaaaatgtt aacagagatt tacacagagg ataattttgt
1141 taagtttttt aaagtactta acagaaaaac atatttgaat tttgataaag ccgtatttaa
1201 gataaatata gtacctaagg taaattacac aatatatgat ggatttaatt taagaaatac
1261 aaatttagca gcaaacttta atggtcaaaa tacagaaatt aataatatga attttactaa
1321 actaaaaaat tttactggat tgtttgaatt ttataagttg ctatgtgtaa gagggataat
1381 aacttctaaa actaaatcat tagataaagg atacaataag gcattaaatg atttatgtat
1441 caaagttaat aattgggact tgtttttttag tccttcagaa gataatttta ctaatgatct
1501 aaataaagga gaagaaatta catctgatac taatatagaa gcagcagaag aaaatattag
1561 tttagattta atacaacaat attatttaac ctttaatttt gataatgaac ctgaaaatat
1621 ttcaatagaa aatctttcaa gtgacattat aggccaatta gaacttatgc ctaatataga
1681 aagatttcct aatggaaaaa agtatgagtt agataaatat actatgttcc attatcttcg
1741 tgctcaagaa tttgaacatg gtaaatctag gattgcttta acaaattctg ttaacgaagc
1801 attattaaat cctagtcgtg tttatacatt tttttcttca gactatgtaa agaaagttaa
```

```
1861 taaagctacg gaggcagcta tgtttttagg ctgggtagaa caattagtat atgattttac
1921 cgatgaaact agcgaagtaa gtactacgga taaaattgcg gatataacta taattattcc
1981 atatatagga cctgctttaa atataggtaa tatgttatat aaagatgatt ttgtaggtgc
2041 tttaatattt tcaggagctg ttattctgtt agaatttata ccagagattg caatacctgt
2101 attaggtact tttgcacttg tatcatatat tgcgaataag gttctaaccg ttcaaacaat
2161 agataatgct ttaagtaaaa gaaatgaaaa atgggatgag gtctataaat atatagtaac
2221 aaattggtta gcaaaggtta atacacagat tgatctaata agaaaaaaaa tgaaagaagc
2281 tttagaaaat caagcagaag caacaaaggc tataataaac tatcagtata atcaatatac
2341 tgaggaagag aaaaataata ttaattttaa tattgatgat ttaagttcga aacttaatga
2401 gtctataaat aaagctatga ttaatataaa taaatttttg aatcaatgct ctgtttcata
2461 tttaatgaat tctatgatcc cttatggtgt taaacggtta gaagattttg atgctagtct
2521 taaagatgca ttattaaagt atatatatga taatagagga actttaattg gtcaagtaga
2581 tagattaaaa gataaagtta ataatacact tagtacagat atacctttтc agctttccaa
2641 atacgtagat aatcaaagat tattatctac atttactgaa tatattaaga atattattaa
2701 tacttctata ttgaatttaa gatatgaaag taatcattta atagacttat ctaggtatgc
2761 atcaaaaata aatattggta gtaaagtaaa ttttgatcca atagataaaa atcaaattca
2821 attatttaat ttagaaagta gtaaaattga ggtaatttta aaaaatgcta ttgtatataa
2881 tagtatgtat gaaaatttta gtactagctt ttggataaga attcctaagt attttaacag
2941 tataagtcta aataatgaat atacaataat aaattgtatg gaaaataatt caggatggaa
3001 agtatcactt aattatggtg aaataatctg gactttacag gatactcagg aaataaaaca
3061 aagagtagtt tttaaataca gtcaaatgat taatatatca gattatataa acagatggat
3121 ttttgtaact atcactaata atagattaaa taactctaaa atttatataa atggaagatt
3181 aatagatcaa aaaccaattt caaatttagg taatattcat gctagtaata atataatgtt
3241 taaattagat ggttgtagag atacacatag atatatttgg ataaaatatt ttaatctttt
3301 tgataaggaa ttaaatgaaa aagaaatcaa agatttatat gataatcaat caaattcagg
3361 tattttaaaa gacttttggg gtgattattt acaatatgat aaaccatact atatgttaaa
3421 tttatatgat ccaaataaat atgtcgatgt aaataatgta ggtattagag gttatatgta
3481 tcttaaaggg cctagaggta gcgtaatgac tacaaacatt tatttaaatt caagtttgta
3541 taggggggaca aaatttatta taaaaaaata tgcttctgga aataaagata atattgttag
3601 aaataatgat cgtgtatata ttaatgtagt agttaaaaat aaagaatata ggttagctac
3661 taatgcatca caggcaggcg tagaaaaaat actaagtgca ttagaaatac ctgatgtagg
3721 aaatctaagt caagtagtag taatgaagtc aaaaaatgat caaggaataa caaataaatg
3781 caaaatgaat ttacaagata ataatgggaa tgatataggc tttataggat ttcatcagtt
3841 taataatata gctaaactag tagcaagtaa ttggtataat agacaaatag aaagatctag
3901 taggactttg ggttgctcat gggaatttat tcctgtagat gatggatggg gagaaaggcc
3961 actgtaatta atctcaaact acatgagtct gtcaagaatt ttctgtaaac atccataaaa
4021 attttaaaat taatatgttt aagaataact agatatgagt attgctatgc taatatctag
4081 ttatttttaat ttattcaata ttattacagt aagaaaaaat actattttta ttgtaaatac
4141 aagtttagtg gtatatctca taaatgatac aagatatcat tataatgatt ttgcaaatta
4201 tagttttgaa taaatatatt tacagtattt ttgaaatgat aataattact tcaaattctt
4261 tagtataatt ttttaatgtc ttaattttta ca
//
```

EXHIBIT 2

Eur. J. Biochem. 189, 73–81 (1990)

The complete amino acid sequence of the *Clostridium botulinum* type A neurotoxin, deduced by nucleotide sequence analysis of the encoding gene

Daphne E. THOMPSON[1], John K. BREHM[1], John D. OULTRAM[1], Tracy-Jane SWINFIELD[1], Clifford C. SHONE[2], Tony ATKINSON[1], Jack MELLING[2] and Nigel P. MINTON[1]

[1] Division of Biotechnology, and [2] Division of Biologics, Public Health Laboratory Service, Centre for Applied Microbiology and Research, Porton Down, England

(Received October 18, 1989) – EJB 89 1261

A 26-mer oligonucleotide probe was synthesized (based on the determined amino acid sequence of the N-terminus of the *Clostridium botulinum* type A neurotoxin, BoNT/A) and used in Southern blot analysis to construct a restriction map of the region of the clostridial genome encompassing BoNT/A. The detailed information obtained enabled the cloning of the structural gene as three distinct fragments, none of which were capable of directing the expression of a toxic molecule. The central portion was cloned as a 2-kb *Pvu*II – *Taq*I fragment and the remaining regions of the light chain and heavy chain as a 2.4-kb *Sca*I – *Taq*I fragment and a 3.4-kb *Hpa*I – *Pvu*II fragment, respectively. The nucleotide sequence of all three fragments was determined and an open reading frame identified, composed of 1296 codons corresponding to a polypeptide of 149 502 Da. The deduced amino acid sequence exhibited 33% similarity to tetanus toxin, with the most highly conserved regions occurring between the N-termini of the respective heavy chains. Conservation of Cys residues flanking the position at which the toxins are cleaved to yield the heavy chain and light chain allowed the tentative identification of those residues which probably form the disulphide bridges linking the two toxin subfragments.

Toxigenic strains of *Clostridium botulinum*, an anaerobic spore-forming bacteria, produce one or more of seven immunologically distinct neurotoxins. They act primarily at the neuromuscular junction, causing paralysis by inhibiting the release of the neurotransmitter acetylcholine [1]. The botulinum neurotoxins are synthesised as a single polypeptide chain but in their most active form comprise of two asymmetric polypeptide units; a light chain (50–59 kDa) and a heavy chain (85–105 kDa), linked by at least one disulphide bridge [2].

Although botulinum neurotoxins exhibit high degrees of similarity (when compared to each other and related tetanus and diphtheria toxins) in their molecular masses and modes of action, they still remain poorly characterised at the molecular level. Proposals for the mechanism of toxic action implicates three separate functional domains. The light chain has the pharmacological activity, while the N-terminal and C-terminal of the heavy chain mediate channel formation and toxin binding, respectively. These heavy-chain, toxin-binding functional domains have been demonstrated with diphtheria toxin [3–5] and with tetanus toxin [6–8]. Channel forming activities have been demonstrated for the N-terminus of the heavy chain with *C. botulinum* type A [9, 10] and there is circumstantial evidence for the association of binding properties with the C-terminus of the heavy chain [10–12].

The entire nucleotide sequence of diphtheria toxin [13, 14] and of tetanus toxin [15, 16] have been determined. Aside from N-terminal and C-terminal sequencing of purified toxin sub-fragments [11, 17, 18], the complete primary amino acid sequence of a botulinum neurotoxin remains undetermined. In the present study we present the entire amino acid sequence of the *C. botulinum* type A neurotoxin (BoNT/A), deduced by nucleotide sequence analysis of the encoding gene. The gene was cloned in *Escherichia coli* as three separate sub-fragments, none of which were capable of directing the expression of a toxinogenic polypeptide.

Correspondence to N. P. Minton, Molecular Genetics Group, Division of Biotechnology, Public Health Laboratory Service, Centre for Applied Microbiology and Research, Porton Down, Salisbury, Wiltshire, SP4 0JG, England

Note. The novel nucleotide sequence data published here has been deposited with the EMBL/GenBank® sequence data banks and is available under accession number X52066.

The novel amino acid sequence data published here has been deposited with the PIR sequence data bank.

Abbreviations. BoNT/A, *Clostridium botulinum* type A neurotoxin.

MATERIALS AND METHODS

Materials

Restriction endonucleases, DNA modifying enzymes and the pUC sequencing kit were obtained from Boehringer Mannheim and used under the conditions recommended by the supplier. [α-^{35}S]dATP, [α-^{32}P]dATP and [γ-^{32}P]rATP and the Multiprime labelling kit were purchased from Amersham International. Zeta Probe and nitrocellulose filters were obtained from Bio-Rad, and from Schleicher & Schull, respectively. Agarose was obtained from Seakem. All oligonucleotides were synthesised on an Applied Biosystems 380A DNA synthesiser by solid-phase synthesis using the phosphite triester method [19].

Strains, vectors and growth conditions

The *E. coli* host used for all recombinant manipulations was TG1 [*Δ(lac-pro)thi supE hsdD5* F′-*traD36 proAB*$^+$ *ΔlacZ*

M15], and the source of the *C. botulinum* type A gene was strain NCTC 2916. The cloning vectors employed were plasmid pMTL23 [20] and the M13 cloning vectors M13mp8 and M13mp9 [21]. *E. coli* was routinely cultured in L broth (1% tryptone, 0.5% yeast extract, 0.5% NaCl), supplemented where appropriate with ampicillin (50 μg/ml). Solidified agar contained 2.0% (mass/vol.) agar (Difco Laboratories). *C. botulinum* was grown in USB II broth as previously described [11].

Nucleic acid manipulations

Standard techniques, including cloning procedures, plasmid preparations, and bacterial transformations have been described elsewhere [22]. Chromosomal DNA was isolated from *C. botulinum* type A NCTC 2916 by the method of Marmur [23]. Cloning of *C. botulinum* type A toxin DNA fragments was performed under GMP II conditions. The generation of M13 templates by the sonication procedure was as previously described [24]. Nucleotide sequencing was undertaken by the dideoxy-chain-termination procedure [25], sequence data being analysed by the computer software of DNASTAR Inc.

Screening procedures

Restriction endonuclease digests of *C. botulinum* type A total DNA were electrophoresed in 0.8% (mass/vol.) agarose in 90 mM Tris/HCl, 90 mM boric acid, 3 mM EDTA, pH 8.3, prior to depurination and transfer of the DNA to the Zeta Probe membrane according to literature procedures [26]. The 26-mer probe was labelled at the 5′-end by transfer of [γ-^{32}P]rATP with T4 polynucleotide kinase [27]. DNA probes, derived from cloned fragments of the BoNT/A gene, were radiolabelled with the Multiprime labelling system according to the manufacturer's instructions. Hybridisation in 0.3 M sodium chloride, 30 mM sodium citrate (NaCl/Cit) using the oligonucleotide probe was undertaken by gradual cooling from 80 °C to 25 °C over a 4 h period. Filters were washed at 55 °C in 0.1% SDS and NaCl/Cit. Hybridisation using BoNT/A sub-fragments was performed overnight at 42 °C, essentially as described [28], except that dextran sulphate was omitted. After hybridisation, filters were dried at 37 °C, mounted on Whatman 3MM paper, covered with Saran Wrap and exposed to X-ray film (Kodak X-Omat) with an intensifying screen for 1 – 7 days at −70 °C. Recombinant clones were screened by *in situ* colony hybridisation [29] for the presence of BoNT/A-specific inserts with a radiolabelled oligonucleotide probe or cloned fragments from the BoNT/A gene. Total cell lysates of positive clones were tested in mice [30] to exclude the possibility of toxic polypeptide expression.

RESULTS AND DISCUSSION

Genomic mapping of the neurotoxin gene

The first 35 amino acids of the N-terminus of the purified 49-kDa fragment (H_2) of BoNT/A has previously been determined [11]. As amino acids 6 – 14 represent the peptide sequence exhibiting minimal translational degeneracy, an encoding 26-mer oligonucleotide was synthesised (Fig. 1) to allow the detection of the gene by DNA · DNA hybridisation techniques. The choice of nucleotide at positions of degeneracy was made both on the basis of the nucleotide present in the equivalent codon of the closely related tetanus gene [15, 16] and on the general codon bias exhibited by clostridial genes [13]. Additionally, in observing the convention that d(G · T) pairing is neutral in terms of base pairing, the strand synthesised (Fig. 1) possessed the maximum number of wobble position dT residues. Southern hybridisation experiments were undertaken using the radiolabelled 26 mer oligonucleotide, as a probe, and genomic DNA isolated from *C. botulinum* type A. The data obtained enabled the assignment of restriction sites to the region of the clostridial genome which encodes the neurotoxin gene (Fig. 2).

Cloning of the central portion of the BoNT/A gene

Data from Southern hybridisation experiments demonstrated that DNA encoding amino acids 6 – 14 of the heavy chain, resided on a 5.0-kb *Pvu*II restriction fragment, which was further divided upon cleavage with *Taq*I into a 3.0-kb and a 2.0-kb *Pvu*II – *Taq*I fragment, the latter of which encoded the N-terminus of the H_2 subunit. Accordingly, genomic DNA from *C. botulinum* type A digested with *Pvu*II was size-fractionated and fragments of around 5.0 kb isolated and subjected to further digestion with *Taq*I. After size-fractionation, restricted DNA of approximately 2.0 kb was purified and shown to contain a restriction fragment capable of hybridising the oligonucleotide probe. The DNA fragments were ligated to *Sma*I/*Acc*I-digested pMTL23 [20] and 1500 recombinant transformants were screened for the presence of type A neurotoxin-specific sequences using *in situ* colony hybridisation and a radiolabelled oligonucleotide probe. A total of five positive clones were obtained, from which one was chosen; its recombinant plasmid designated pCBA2. Restriction enzyme analysis of the insert of pCBA2 demonstrated the presence of a DNA insert of the expected size.

By following the above cloning strategy, the chances of generating an *E. coli* recombinant clone capable of producing a toxigenic molecule were negligible. This conclusion was based on a number of factors. Principal amongst these are the observations [11] that both the heavy chain and light chain are required for toxicity (i.e. purified preparations of either subunit are non-toxic), removal of the C-terminus of the toxin results in an inactive molecule (composed to the entire light chain and 54% of the heavy chain) and the large size of the structural gene (predicted to be greater than 4.0 kb). Thus on size considerations, the 2.0-kb *Pvu*II – *Taq*I fragment could not encode a toxic molecule. Furthermore, the method used in purifying DNA fragments to be cloned eliminated the risk of concomitant cloning the contiguous regions of the gene, i.e. the larger portion of the 5.0-kb *Pvu*II fragment could not have been isolated following cleavage with *Taq*I and subsequent size fractionation. Nevertheless, toxicity tests were routinely performed on the lysates from all primary clones to exclude the possibility that a toxic molecule was being produced.

The cloning vector pMTL23 was specifically designed to facilitate the generation of M13 templates by the sonication procedure [20]. Cleavage of pCBA2 with the restriction enzymes *Bam*HI and *Bgl*II allowed the excision of the clostridial DNA insert as a 2.05-kb DNA fragment with compatible cohesive termini. The gel-purified DNA was circularised by self-ligation, fragmented by sonication and the random DNA strands generated as blunt-ended fragments by treatment with T4 polymerase and cloned into the *Sma*I site of M13mp8. The nucleotide sequence data obtained from 200 such recombinant templates was compiled into a single contiguous sequence using the computer software of DNASTAR INC. The insert

```
 C   I   K   I   K   N   E   D   L
TGT ATA AAA ATT AAA AAT GAA GAT TT    - tetanus gene sequence

 C   I   K   V   N   N   W   D   L
TGT ATA AAA GTT AAT AAT TGG GAT TT    - botulinum probe
||| ||  ||| ||| ||| ||| ||| ||: ||
ACA TAg TTT CAA TTA TTA ACC CTG AA    - actual sequence
```

Fig. 1. *Selection of an oligonucleotide gene probe for the BoNT/A gene.* The nucleotides incorporated at positions of degeneracy in the oligonucleotide probe (bold) were chosen on the basis of clostridial codon usage [31] and those present in the equivalent position in the tetanus gene [16]. The actual sequence of the BoNT/A gene is shown below the oligonucleotide probe, identity between the two being indicated by |, mismatch by a lower case letter in the actual sequence, and the G:T pairing by a colon. Actual sequence refers to the BoNT/A gene sequence

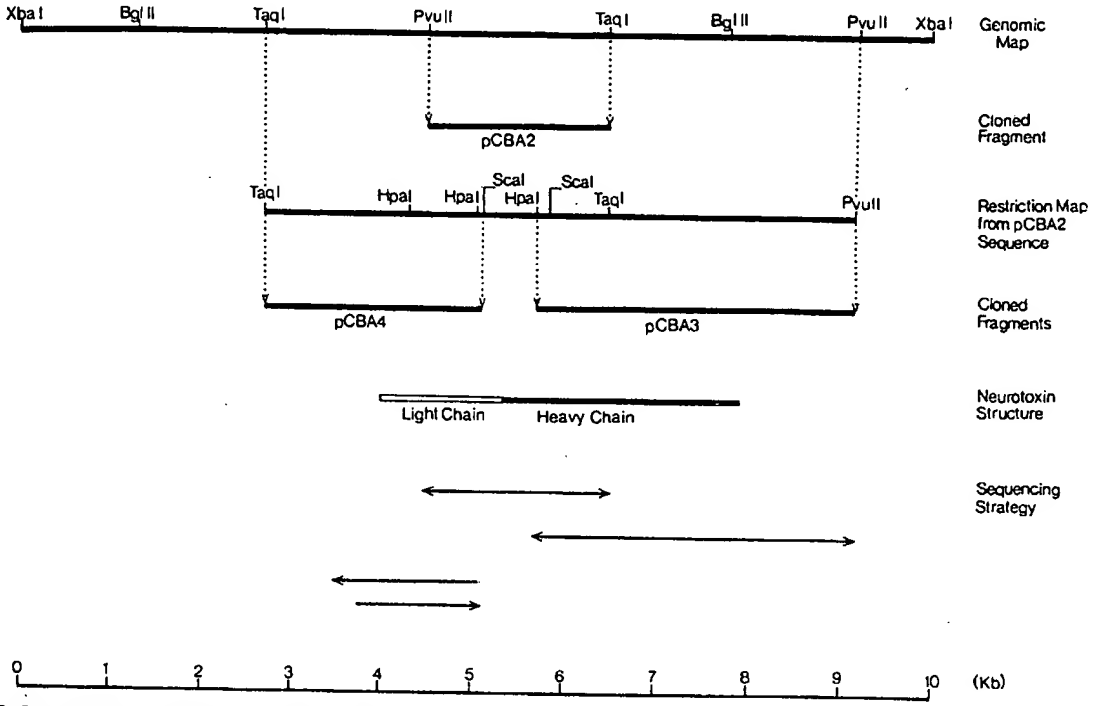


Fig. 2. *BoNT/A gene cloning and sequencing strategy.* The restriction map of the *C. botulinum* genomic region encoding BoNT/A, depicted at the top of the diagram, was derived by Southern blot analysis of NCTC 2916 chromosomal DNA using the 26-mer oligonucleotide as a radiolabelled probe. The respective regions of the genome carried by the recombinant plasmids pCBA2, pCBA3 and pCBA4, and the proportion of the BoNT/A gene which they encode, are indicated below this map. Regions encompassed by double-headed arrows (the inserts carried by pCBA3 and pCBA2) were sequenced by the sonication procedure [24]. The region of the pCBA4 insert encompassed by the single-headed arrows was sequenced by primer extension in the appropriate direction

of pCBA2 proved to be 1845 bp in length (Fig. 1), and in common with previously cloned clostridial DNA, had a low G/C content (27.53%).

Translational of the DNA in the six possible reading frames demonstrated the presence of a single open reading frame (622 codons) encompassing the entire cloned region. With a single discrepancy (see below), the presence of an amino acid sequence corresponding to that derived from the purified H_2 fragment [11] confirmed that this open reading frame encoded BoNT/A. Examination of the nucleotide sequence in this region indicated that of the eight potential mismatches between the oligonucleotide probe and the antisense DNA strand, only two had occurred, one of which involved a neutral d(G · T) pairing (Fig. 1). From comparison of the amino acid sequence with tetanus toxin, it was estimated that pCBA2 encoded 50% of the BoNT/A light chain and 38% of the heavy chain.

Having cloned the central portion of the BoNT/A gene, specific sub-fragments of the pCBA2 insert were used as radiolabelled probes to identify, and subsequently clone, genomic fragments encoding the remainder of the gene.

Cloning of the 3'-end of the BoNT/A gene

A 650-bp *Hpa*I – *Taq*I restriction fragment was isolated from pCBA2 and used as the radiolabelled probe in a Southern blot analysis of *C. botulinum* type A genomic DNA. Restriction enzymes utilised in the cleavage of chromosomal DNA were those known to cleave both the internal (identified from the nucleotide sequence) and external (using the genomic map data) cloned insert of pCBA2. The data obtained indicated that the 3'-end of the gene resided on a 3.4-kb *Hpa*I – *Pvu*II fragment. Accordingly, genomic DNA was cleaved with *Hpa*I and *Taq*I, size-fractionated, and DNA fragments of approxi-

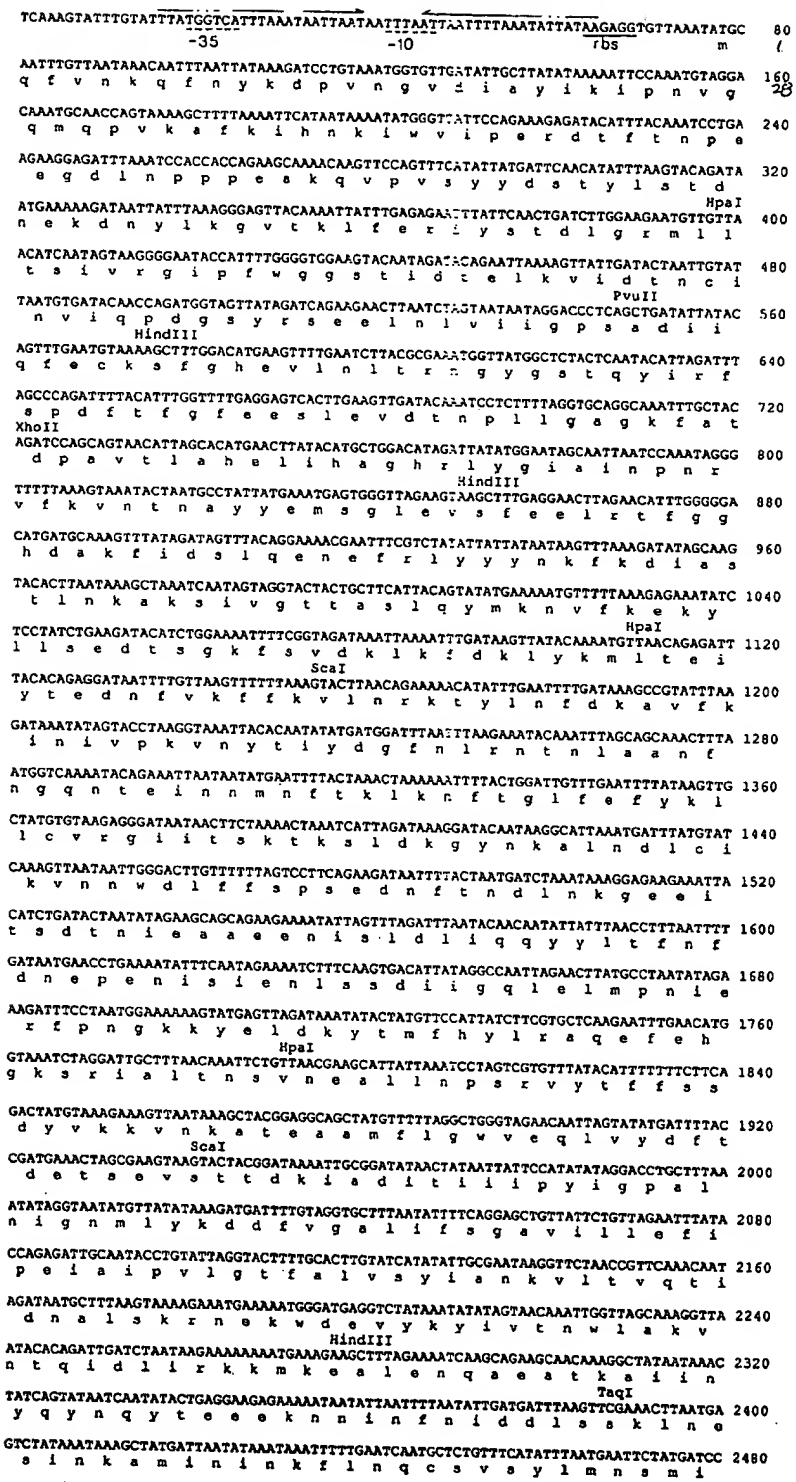
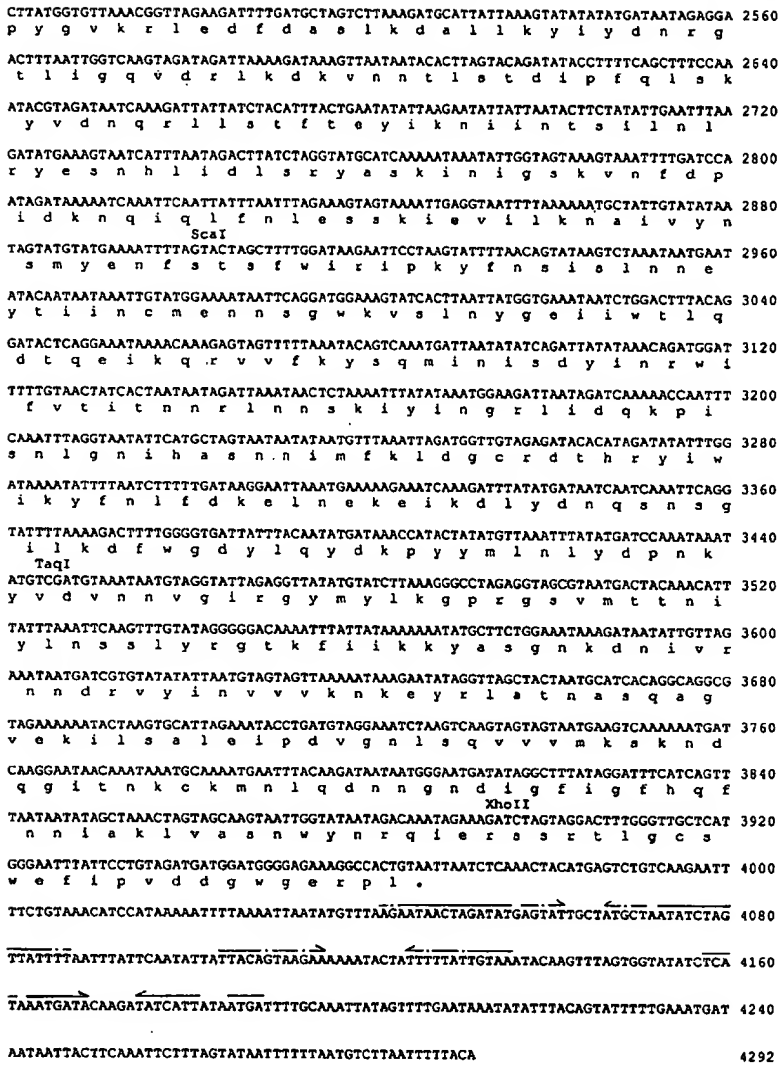


Fig. 3. *Nucleotide sequence of the BoNT/A gene.* The illustrated sequence represents a 4292-base portion of the *C. botulinum* NCTC 2916 genome encoding BoNT/A derived from the sequences of the recombinant plasmids pCBA2, pCBA3 and pCBA4. The amino acid sequence of BoNT/A is given below the second nucleotide of the appropriate codon in single-letter code. The putative −35 and −10 sequences are marked by dashed lines below the sequence, the ribosome-binding sites by underlining of the relevant sequence, and regions of dyad symmetry by facing arrows above the sequence. In the latter instance a gap in the overlining represents mismatch, while the presence of a dot indicates a d(G · T) pairing



mately 3.4 kb in size isolated and cloned into pMTL23 between the *Sma*I and *Cla*I sites. It should be noted that the cloning strategy ensured that concomitant cloning of contiguous regions of the BoNT/A gene was extremely remote. Although cleavage of genomic DNA with both *Hpa*I and *Pvu*I releases the 3'-end of the gene as the desired 3.4-kb restriction fragment, the adjacent 5'-end of the gene is fragmented into a 0.5-kb *Hpa*I, a 0.9-kb *Hpa*I – *Pvu*II and a 0.25-kb *Pvu*II – *Hpa*I sub-fragment. Such fragments would not be purified during the isolation of the 3.4-kb DNA restriction fragments.

A total of 1500 recombinant clones were screened by *in situ* colony hybridisation, using the insert of pCBA2 as a radiolabelled probe, and five positive clones identified. One such clone was chosen for further analysis and its plasmid isolated (designated pCBA3), and shown to contain a DNA insert of the expected size. The nucleotide sequence of the insert of pCBA3 was determined exactly as described for pCBA2. The sequence obtained demonstrated the expected overlap with that of pCBA2, allowing the identification of the translational stop codon of the open reading frame presumed to encode BoNT/A (see below).

Cloning of the 5'-end of the BoNT/A gene

The cloning of the 5'-end of the BoNT/A gene was undertaken in a manner analogous to that described for the 3'-end. In this case the DNA fragment used as a radiolabelled probe in a Southern blot analysis of the *C. botulinum* type A genome was a 0.9-kb *Pvu*II – *Hpa*I fragment derived from pCBA2. The restriction fragment targeted for cloning was a 2.4-kb *Sca*I – *Taq*I fragment. Concomitant cloning of adjacent BoNT/A sequences was again avoided since cleavage of chromosomal DNA with *Sca*I and *Taq*I results in the fragmentation of the central portion of the gene into a 0.55-kb *Taq*I – *Sca*I fragment and two *Sca*I restriction fragments of 0.55 kb and 0.75 kb. Appropriately cleaved genomic DNA was therefore size fractionated and purified DNA fragments of approximately 2.4-kb inserted into pMTL23 cut with *Sma*I and *Cla*I.

A single recombinant plasmid was identified carrying the desired insert and designated pCBA4.

In contrast to the inserts of the previous two clones, the nucleotide sequence of the pCBA4 insert was not derived by the sonication procedure, but was cloned into M13mp9 as a 2.4-kb *Bam*HI–*Bgl*II fragment. The nucleotide sequence of the antisense strand was determined by primer extension using custom synthesised oligonucleotides. As it proved impossible to clone the same fragment into the vector M13mp8, the nucleotide sequence of the sense strand was derived using pCBA4 DNA and the appropriate primers. In both cases the region sequenced corresponded to the structural gene and the immediate 5′-non-coding region.

Features of the coding region

The nucleotide sequences derived from pCBA2, pCBA3, and pCBA4 were compiled into a single contiguous sequence using the computer software of DNASTAR INC. The sequence depicted in Fig. 3 represents a 4292-bp portion of this sequence and has been determined in its entirety on both DNA strands. Translation of the sequence identified an open reading frame of 3891 bp commencing with an AUG start codon and terminating with a UAA stop codon. The deduced polypeptide is composed of 1296 amino acid residues with a predicted molecular mass of 149502 Da. This is in close agreement with the determined molecular mass of purified BoNT/A [11]. Comparison of the predicted amino acid sequence of the open reading frame with published amino acid sequences determined from purified toxin fragments confirmed the gene identified as encoding BoNT/A. Thus the 17 N-terminal amino acid residues encoded by the open reading frame identified agree with the previously determined amino acid sequence of the BoNT/A light chain, except at amino acid position 2 where we have Gln rather than the previously reported Pro [17]. Similarly, with one exception, amino acids 449–465 and 449–483 of our sequence agree with the primary structure determinations undertaken on the purified heavy chain by Sathymoorthy and co-workers [18] and Shone and co-workers [11], respectively. The one discrepancy arises at position 480 where we predict Glu in contrast to the previously reported Pro [11]. The sequence of the predicted polypeptide of amino acids 873–896 exhibit a high degree of similarity to the determined amino acid sequence of the BoNT/A C-terminus [18]. The two exceptions are amino acid positions 876 and 892, where the predicted residues are Thr876 and Ser896, which replace the previously determined Leu and Lys residues, respectively. It is possible that either the small amounts of protein available for analysis led to sequence-determination errors or that strain differences may exist.

The codon usage exhibited by the BoNT/A gene conforms to the pattern generally seen for genes isolated from *Clostridium species* whose DNA displays a high (70%) d(A/T) content [31–35]. The principal features of this pattern are the use of AUG and UAA as the respective translational initiation and termination codons, and a strong discrimination against all degenerate codons ending in C or G or, in the case of Ser and Arg, beginning with C. Thus in the BoNT/A gene, 86.1% of Arg codons conform to AGN rather than CGN, 69% of Leu codons conform to UUA as opposed to CUN, while overall, 90.3% of the degenerate codons end in A or U. The one exception to this rule appears to be in the choice of Lys codons, where the frequency of occurrence of the codons AAA and AAG is almost equal, being 24 and 20 respectively. This lack of bias is in contradiction to other clostridial genes, e.g. the distribution in the related tetanus gene is 98 AAA codons versus 9 AAG codons. A consequence of the observed codon bias is that many of those codons known to act as modulators of gene expression in *E. coli* occur frequently in clostridial genes [31]. The BoNT/A gene therefore exhibits a 53.8% preference for the AUA (Ile) codon, 43.7% preference for the GGA (Gly) codon and an overall 86.1% preference for the AGN (Arg) codon. Other modulator codons [CGA (Arg), CGC (Arg) and CUA (Leu)] are used infrequently.

With exception of the N-terminal Met, the deduced amino acid sequence of BoNT/A corresponds exactly to that determined experimentally from purified toxin. Thus, as is the case with tetanus toxin [16], secretion of BoNT/A toxin is not mediated by possession of a signal peptide [36]. Post-translational modification of the nascent polypeptide N-terminus is therefore limited to removal of a methionine residue.

Features of noncoding region

The noncoding region on the 5′-position adjacent to the BoNT/A structural gene was examined for regulatory sequences. Putative ribosome binding sites [37] were identified five and eight nucleotides upstream from the AUG start codon, d(AGGTGT) and d(AAGAGG), respectively (Fig. 2). The latter sequence would appear to be the more favourable candidate for the BoNT/A ribosome-binding site, since (a) the average spatial distance between the translational start codon and the ribosome-binding site sequences of clostridial genes is eight nucleotides [31]; (b) the BoNT/A ribosome-binding site exhibits greater overall similarity with procaryotic 16S rRNA [38]; (c) the alternative candidate, d(AAGTGT), is preceded by dG, a factor associated with misalignment [39].

A putative promoter region, similar to *E. coli* and *Bacillus subtilis* promoters [40], was identified [d(-35-TGGTCA); d(-10-TTTAAT)] 5′ to the BoNT/A structural gene. The spacial distance between these two sequences (16 nucleotides) is consistent with that found in *E. coli* [40], *B. subtilis* [41] and *Clostridium* spp. [31] viz., ± 17, ± 16–19 and ± 16 or ± 17, respectively. Further analysis of the region for the extended promoter sequence found in Gram-positive organisms [42], showed that the highly conserved dT residue at −16, the dA residue at −6 and the dT residues at −5 and −3 occurred. Only 7 of the 12 conserved residues in the −4 to −18 region were observed [42], however this may be a consequence of the high d(A/T) content of the DNA (82%) in the noncoding region.

The AUG translational initiation codon of the BoNT/A gene is preceded by a region of dyad symmetry, which if transcribed would form a hairpin loop structure with ΔG = −48.5 kJ [43]. In such a structure the putative −35 region would be concealed within the stem (as would the extreme 5′-end of the ribosome-binding site), while the −10 sequence would be situated within the loop. Stem loops 5′ to the ribosome-binding site and start codon have been described in *E. coli*, and implicated in mRNA stability by protecting the translation initiation complex [44]. A second region of dyad symmetry was observed 75 nucleotides 3′ to the UAA termination codon, and resembles a ρ-independent transcriptional terminator since it is followed by a stretch of dT (five out of eight). A mRNA transcript of this region would have a calculated ΔG = −71.1 kJ [43]. Two further regions of dyad symmetry were also present on the 3′ side of this sequence, although the stem loop structures that could be formed were thermodynamically less stable, having ΔG values of −19.2 kJ

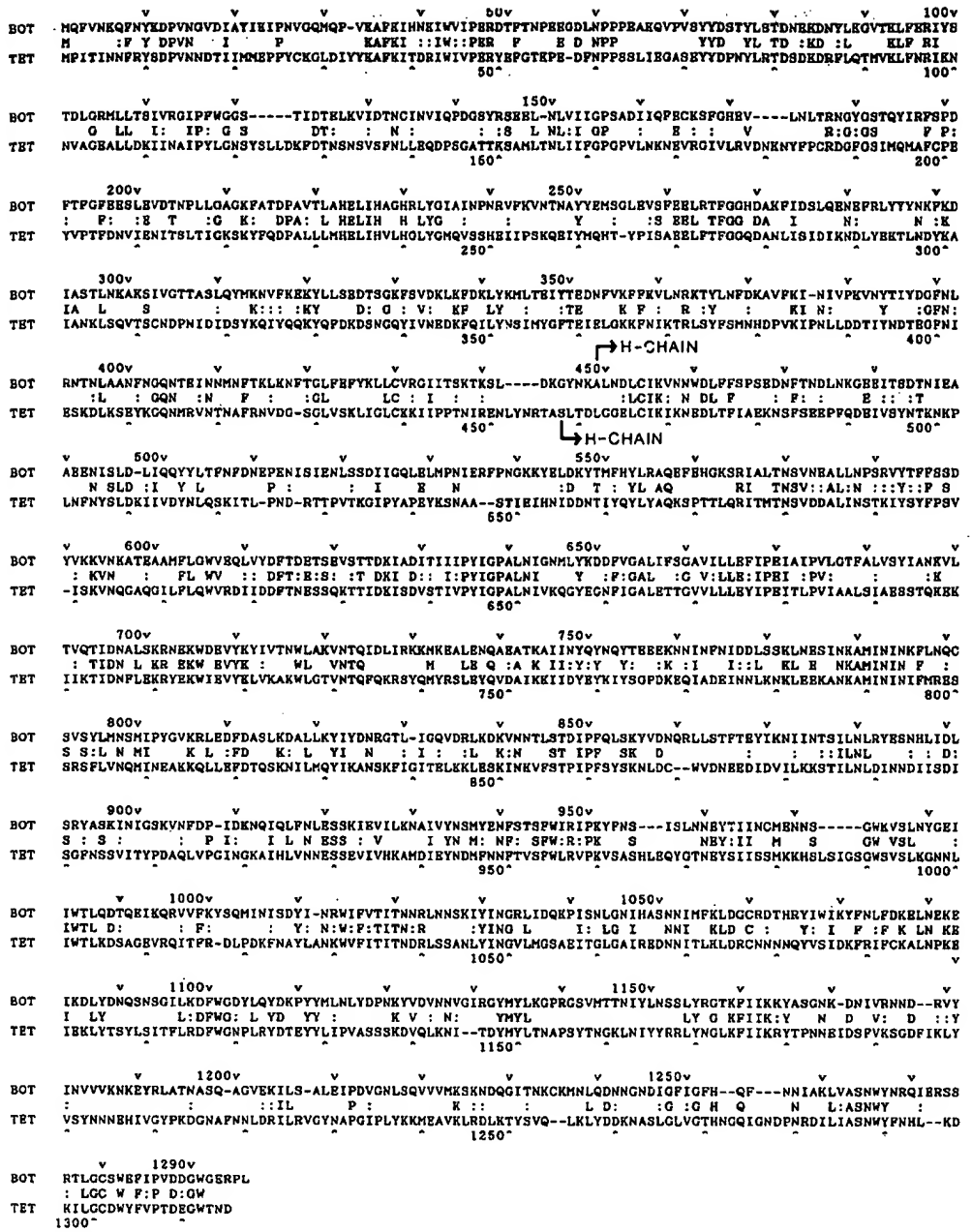


Fig. 4. *Comparative alignment of the primary amino acid sequences of the BoNT/A (BOT) and tetanus (TET) neurotoxins.* Identical amino acids are as shown, conservative replacements are indicated by colons. The N-terminal amino acids of the heavy chain (H-CHAIN) of both toxins are marked by right-angled arrows. The Cys residues thought to be involved in the formation of the disulphide bridge, linking the heavy chain and light chain, are the first Cys residues on either side of these arrows

(nucleotides 4104–4136) and $\Delta G = -34.3$ kJ (nucleotides 4158–4188).

Similarity to tetanus toxin

The structural and pharmacological relatedness of botulinum and tetanus toxin has long been recognised. Although the recent determination of the amino acid sequence of tetanus toxin allowed limited comparisons with botulinum toxins to be made [16], the availability of the entire amino acid sequence of BoNT A now allows a more complete comparative analysis. Comparative alignment of tetanus toxin and BoNT/A neurotoxin (Fig. 3) reveals that the two toxins exhibit 32.9% similarity. It is apparent that overall the heavy chains exhibit a higher degree of similarity than the light chains, at both the level of identity (heavy chain, 34.5%; light chain 29.5%) and

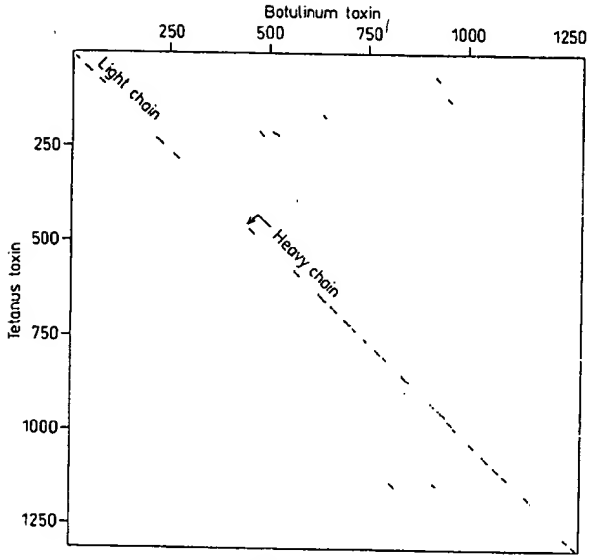


Fig. 5. *Dot matrix plot of BoNT/A and tetanus toxin similarity.* Similarity parameters show a 60% match with a window size of 10 amino acids

similarity (heavy chain, 47.9%; light chain 41.8%). This is expected since the heavy chain is the receptor-binding domain and the light chain the effector, and such membrane/receptor-binding proteins have similar structural requirements. It is apparent that specific regions of both the heavy chains and light chains exhibit substantial similarity. The higher degree of conservation of primary sequence between the heavy chains of the two toxins is more apparent when the two proteins are aligned using a dot matrix plot. The illustration in Fig. 4 represents such an alignment, using a window size of 10 amino acids and scoring 60% similarity.

Conservation of Cys residues in tetanus and botulinum toxins was observed at positions 1060 and 1280. Of particular note is the conservation of Cys454 which occurs at the same position in the N-termini of *C. botulinum* type A, B and E and on the tetanus toxins [16]. Cys454 is the only Cys residue in the N-terminal region of the heavy chain and is probably involved in the disulphide bridge to the light chain. Similarly, Cys430 on the light chain of the botulinum toxin is conserved in the same position as in the tetanus toxin, and from the positions and predicted structural comparisons of the other two Cys residues in this domain, is the most probable Cys residue to form a disulphide bridge linking the light and heavy chains.

The similarity shared by the botulinum type A and tetanus toxins, and pattern of alignment displayed in the dot-matrix plot (Fig. 3), suggest that they are derived from a common ancestral gene. Similar conclusions have been drawn from comparisons of short N-terminal amino acid sequences of the light chains and heavy chains of purified tetanus and botulinum type A, B and E toxin fragments [16].

Structural features

Comparative alignment of tetanus and BoNT/A with a dot-matrix plot identified six regions where 80% similarity occurred (regions in question are marked on Fig. 3). Of these regions, four occurred in the N-terminus of the heavy chain. Hydrophilicity plots of the predicted amino acid sequence of botulinum type A toxin, according to the methods of Kyte and Doolittle [45], and Hopp and Woods [46], showed that three of these highly conserved sequences occurred in essentially hydrophobic regions. The first of these predicted regions contained 19 residues of uncharged amino acids from Ile630–Tyr648. The second region, which is separated from the first region by three charged amino acid residues, extends from Phe652–Asn687 and contains two acidic residues. The third region involves Leu773–Val807. The ratio of hydrophobic amino acids/polar amino acids contained in the first two regions was greater (74:20) than the third region (54:40). Hydrophilicity plots [45, 46] of tetanus toxin showed that the equivalent three regions also occurred in areas of hydrophobicity. One of these regions in tetanus toxin (Asn660–Ala691) has been identified as sufficient in length and hydrophobicity to potentially span mammalian cell membranes [16]. The first two hydrophobic regions identified in botulinum type A toxin also have this potential. The conserved regions occurring in the N-terminal domain of the heavy chain are hydrophobic, which is implicated in channel-forming activities in both toxins *in vitro* [6, 7, 9, 11]. Although diphtheria toxin is structurally and functionally similar to the botulinum and tetanus neurotoxins, no significant overall similarity to the *C. botulinum* type A toxin was found.

Of the three known proteolytic cleavage sites in the botulinum toxin protein sequence, the first (448–449), which produces toxin from the inactive precursor, occurs within a predicted short helix between two areas of β-pleated sheets. The papain cleavage site (856–857) occurs at a β-turn between two β-pleated sheets. The tryptic cleavage site (873–874), which is specific to botulinum toxin, again occurs at a β-turn between two β-pleated sheets.

Although amino acid sequence fingerprinting [47] for the occurrence of the classical ADP ribose binding fold [48] reveals no sign of this motif, the light chain of BoNT/A has a predicted structure rich in β-pleated sheets, α-helices and β-pleated sheet folds. In addition the distribution of Lys and Arg residues in the light chain, which predominate the C-terminal one third of the light chain, together with other partial amino acid sequence similarity, suggest a structure capable of binding a nucleic acid or nucleotide cofactor. The closest light chain primary structure similarity is found with a plethora of unrelated proteins but amongst these are the group of acetyl-CoA acyltransferases. More detailed analysis of the strikingly similar amino acid sequences, e.g. 223–230, also reveals an amino acid sequence corresponding to RNA-, DNA- and nucleotide-binding proteins. The Lys and Arg distribution is such that while the whole molecule has a predicted pI of 6.3, the light chain is basic (in contrast to tetanus toxin) and the heavy chain acidic. The predicted secondary structure of the heavy chain is reminiscent of many receptor- or immunoglobulin-binding proteins in that it contains long stretches of α-helix, β-pleated sheet and β-turn structures, and importantly, such structures abound in the C-terminal domain from residue 850 onwards. Comparison of the heavy chain sequences with the data base indicates that the closest primary structure similarity is found with a number of receptors and receptor-binding proteins including the acetyl cholinesterase precursor protein. Secondary structure comparisons of the botulinum and tetanus toxin sequences show greater differences between the light chain structures than with the heavy chain structures, in agreement with the above observations and known functions of these molecules.

Conclusion

C. botulinum is frequently used as a test organism by microbiologists in the food industry, while its neurotoxins are used by a growing number of neurobiochemists in the study of nerve action. More importantly, this neurotoxin is finding ever increasing clinical uses as an alternative to surgical manipulation of a wide range of aberrant muscular functions [49]. A consequence of these applications is the required immunisation of the personnel involved. The availability of the BoNT/A gene sequence will therefore not only facilitate structure/function studies but also allow the production of toxin for clinical use and toxoid for the formulation of improved vaccines.

We are grateful to Ken Fantom and Roy Hartwell for synthesis of the oligonucleotides used in this study and to Nicola Minion for the preparation of this manuscript.

REFERENCES

1. Sugiyama, H. (1980) *Microbiol. Rev. 44*, 419–448.
2. Dasgupta, B. R. & Sugiyama, H. (1972) *Biochem. Biophys. Res. Commun. 48*, 108–112.
3. Chung, D. W. & Collier, R. J. (1977) *Biochim. Biophys. Acta 483*, 248–257.
4. Kagan, B. L., Finkelstein, A. & Colombini, M. (1981) *Proc. Natl Acad. Sci. USA 78*, 4950–4954.
5. Giannini, G., Rappuoli, R. & Ratti, G. (1984) *Nucleic Acids Res. 12*, 4063–4066.
6. Boquet, P. & Duflot, E. (1982) *Proc. Natl Acad. Sci. USA 79*, 7614–7618.
7. Hoch, D. H., Romero-Mira, M., Ehrlich, B. E., Finkelstein, A., Dasgupta, B. R. & Simpson, L. L. (1985) *Proc. Natl Acad. Sci. USA 82*, 1692–1696.
8. Morris, N. P., Consiglio, E., Kohn, L. D., Habig, W. H., Hardgree, M. C. & Heltings, T. B. (1980) *J. Biol. Chem. 255*, 6071–6076.
9. Shone, C. C., Hambleton, P. & Melling, J. (1987) *Eur. J. Biochem. 167*, 175–180.
10. Blaustein, R. O., Germann, W. J., Finkelstein, A. & Dasgupta, B. R. (1987) *FEMS Microbiol. Lett. 226*, 115–120.
11. Shone, C. C., Hambleton, P. & Melling, J. (1985) *Eur. J. Biochem. 151*, 75–82.
12. Maisey, E. A., Wadsworth, J. D. F., Poulain, B., Shone, C. C., Melling, J., Gibbs, P., Taue, L. & Dolly, J. O. (1988) *Eur. J. Biochem. 177*, 683–691.
13. Greenfield, L., Bjorn, M. J., Horn, G., Fong, D., Buck, G. A., Collier, R. J. & Kaplan, D. A. (1983) *Proc. Natl Acad. Sci. USA 80*, 6853–6857.
14. Rappuoli, R., Ratti, G. & Giannini, G. (1983) *Nucleic Acids Res. 11*, 6589–6595.
15. Fairweather, N. F. & Lyness, V. (1986) *Nucleic Acids Res. 14*, 7809–7812.
16. Eisel, U., Jarausch, W., Goretzki, K., Henschen, A., Engels, J., Weller, U., Hudel, M., Habermann, E. & Niemann, H. (1986) *EMBO J. 5*, 2495–2502.
17. Schmidt, J. J., Sartymoorthy, V. & Dasgupta, B. R. (1984) *Biochem. Biophys. Res. Commun. 119*, 900–904.
18. Sathymoorthy, V., Dasgupta, B. R., Foley, J. & Niece, R. L. (1988) *Arch. Biochem. Biophys. 266*, 142–151.
19. Atkinson, T. & Smith, M. (1984) in *Oligonucleotide synthesis: A practical approach* (Gait, M. J., ed.) pp. 88–114, IRL Press, Oxford.
20. Chambers, S. P., Prior, S. E., Barstow, D. A. & Minton, N. P. (1988) *Gene 68*, 139–149.
21. Messing, J. & Vieira, J. (1982) *Gene 19*, 269–276.
22. Minton, N. P., Atkinson, T., Bruton, C. J. & Sherwood, R. F. (1984) *Gene 31*, 31–38.
23. Marmur, J. (1961) *J. Mol. Biol. 3*, 208–218.
24. Minton, N. P., Bullman, H. M. S., Scawen, M. D., Atkinson, T. & Gilbert, H. J. (1986) *Gene 46*, 25–35.
25. Sanger, F., Coulson, A. R., Barrell, B. G., Smith, A. J. H. & Roe, B. A. (1980) *J. Mol. Biol. 143*, 161–178.
26. Reed, K. C. & Mann, D. A. (1985) *Nucleic Acids Res. 20*, 7207–7221.
27. Maxam, A. M. & Gilbert, W. (1977) *Proc. Natl Acad. Sci. USA 74*, 560–564.
28. Hatheway, C. L. (1976) *Appl. Environ. Microbiol. 31*, 234–242.
29. Sager, R., Anisowicz, A. & Hocell, N. (1981) *Cell 23*, 41–50.
30. Grunstein, M. & Hogness, D. (1975) *Proc. Natl Acad. Sci. USA 72*, 3961–3965.
31. Young, M., Minton, N. P. & Staudenbauer, W. L. (1989) *FEMS Microbiol. Rev. 63*, 301–326.
32. Roggentin, P., Rothe, B., Lottspeich, F. & Schauer, R. (1988) *FEBS Lett. 238*, 31–34.
33. Steffen, C. & Matzura, H. (1989) *Gene 75*, 349–354.
34. Iwanejko, L. A., Routledge, M. N. & Stewart, G. S. A. B. (1989) *J. Gen. Microbiol. 135*, 903–909.
35. Leslie, D., Fairweather, N., Pickard, D., Dougan, G. & Kehoe, M. (1989) *Mol. Microbiol. 3*, 383–392.
36. Perlman, D. & Halvorson, H. O. (1983) *J. Mol. Biol. 167*, 391–409.
37. Shine, J. & Dalgarno, L. (1974) *Proc. Natl Acad. Sci. USA 71*, 1342–1346.
38. Van Kippenberg, P. H., Van Kimmenade, J. M. A. & Heus, H. A. (1984) *Nucleic Acids Res. 12*, 2595–2603.
39. Stormo, G. D., Schneider, T. D. & Gold, L. (1982) *Nucleic Acids Res. 10*, 2971–2996.
40. Rosenberg, M. & Court, D. (1979) *Annu. Rev. Genet. 13*, 319–353.
41. Losick, R. & Pero, J. (1981) *Cell 25*, 582–584.
42. Graves, M. C., Mullenbach, G. T. & Rabinowitz, J. C. (1985) *Proc. Natl Acad. Sci. USA 82*, 1653–1657.
43. Tinoco, I., Borer, P. N., Dengler, B., Levine, M. D., Uhlenbeck, O. C., Crothers, D. M. & Gralla, J. (1973) *Nature New Biol. 246*, 40–41.
44. Shinelding, S., Gayle, M., Pribnow, D. & Gold, L. (1987) *Mol. Gen. Genet. 207*, 224–232.
45. Kyte, J. & Doolittle, R. F. (1982) *J. Mol. Biol. 157*, 105–132.
46. Hopp, T. P. & Woods, K. R. (1981) *Proc. Natl Acad. Sci. USA 78*, 3824–3828.
47. Wierenga, R. K., Terpstra, P. & Hol, W. G. J. (1986) *J. Mol. Biol. 187*, 101–107.
48. Rossman, M. G., Moras, D. & Olsen, K. W. (1974) *Nature 259*, 194–199.
49. Shone, C. C. & Hambleton, P. (1989) in *Clostridia* (Minton, N. P. & Clarke, D. J., eds) pp. 265–292, Plenum Publishing Co., New York.

Sequence Revision History

Show difference between I and II as GenBank/GenPept

Entrez

Search for Genes
LocusLink provides curated information for human, fruit fly, mouse, rat, and zebrafish

Help|FAQ

Batch Entrez: Upload a file of GI or accession numbers to retrieve protein or nucleotide sequences

Check sequence revision history

How to create WWW links to Entrez

LinkOut

Cubby

Related resources

BLAST

Reference sequence project

LocusLink

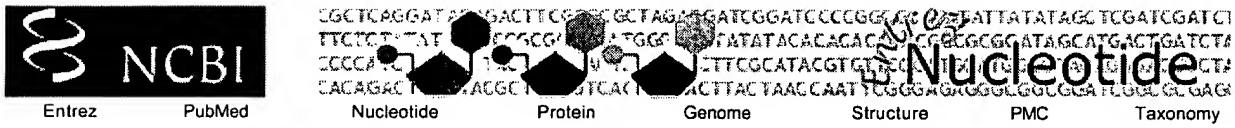
Clusters of orthologous groups

Protein reviews on the web

Revision history for M81186

GI	Version	Update Date	Status	I	II
144734	1	Oct 2 1994 10:20	Live		
144734	1	Apr 26 1993 4:35	Dead		

Accession M81186 was first seen at NCBI on Apr 26 1993 4:35



1: M81186. Clostridium botul...[gi:144734]

Links

```
LOCUS       CLOBOTB                 4041 bp    DNA     linear   BCT 26-APR-1993
DEFINITION  Clostridium botulinum neurotoxin type B (botB) gene, complete cds.
ACCESSION   M81186
VERSION     M81186.1  GI:144734
KEYWORDS    botB gene; neurotoxin type B.
SOURCE      Clostridium botulinum
  ORGANISM  Clostridium botulinum
            Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
            Clostridium.
REFERENCE   1  (bases 1 to 4041)
  AUTHORS   Whelan,S.M., Elmore,M.J., Bodsworth,N.J., Brehm,J.K., Atkinson,T.
            and Minton,N.P.
  TITLE     Complete nucleotide sequence of the Clostridium botulinum gene
            encoding the type B neurotoxin
  JOURNAL   Unpublished (1991)
COMMENT     Original source text: Clostridium botulinum DNA.
FEATURES             Location/Qualifiers
     source          1..4041
                     /organism="Clostridium botulinum"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:1491"
     gene            57..3932
                     /gene="botB"
     CDS             57..3932
                     /gene="botB"
                     /function="vertebrate neurotoxin"
                     /codon_start=1
                     /transl_table=11
                     /product="neurotoxin type B"
                     /protein_id="AAA23211.1"
                     /db_xref="GI:144735"
                     /translation="MPVTINNFNYNDPIDNNNIIMMEPPFARGTGRYYKAFKITDRIW
                     IIPERYTFGYKPEDFNKSSGIFNRDVCEYYDPDYLNTNDKKNIFLQTMIKLFNRIKSK
                     PLGEKLLEMIINGIPYLGDRRVPLEEFNTNIASVTVNKLISNPGEVERKKGIFANLII
                     FGPGPVLNENETIDIGIQNHFASREGFGGIMQMKFCPEYVSVFNNVQENKGASIFNRR
                     GYFSDPALILMHELIHVLHGLYGIKVDDLPIVPNEKKFFMQSTDAIQAEELYTFGGQD
                     PSIITPSTDKSIYDKVLQNFRGIVDRLNKVLVCISDPNININIYKNKFKDKYKFVEDS
                     EGKYSIDVESFDKLYKSLMFGFTETNIAENYKIKTRASYFSDSLPPVKIKNLLDNEIY
                     TIEEGFNISDKDMEKEYRGQNKAINKQAYEEISKEHLAVYKIQMCKSVKAPGICIDVD
                     NEDLFFIADKNSFSDDLSKNERIEYNTQSNYIENDFPINELILDTDLISKIELPSENT
                     ESLTDFNVDVPVYEKQPAIKKIFTDENTIFQYLYSQTFPLDIRDISLTSSFDDALLFS
                     NKVYSFFSMDYIKTANKVVEAGLFAGWVKQIVNDFVIEANKSNTMDKIADISLIVPYI
                     GLALNVGNETAKGNFENAFEIAGASILLEFIPELLIPVVGAFLLESYIDNKNKIIKTI
                     DNALTKRNEKWSDMYGLIVAQWLSTVNTQFYTIKEGMYKALNYQAQALEEIIKYRYNI
                     YSEKEKSNINIDFNDINSKLNEGINQAIDNINNFINGCSVSYLMKKMIPLAVEKLLDF
```

DNTLKKNLLNYIDENKLYLIGSAEYEKSKVNKYLKTIMPFDLSIYTNDTILIEMFNKY
NSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSANSKIRVTQN
QNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKISIRGNRII
WTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESNTDIKD
IREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLKDFW
GNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKFI
IRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDEF
YNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCISKW
YLKEVKRKPYNLKLGCNWQFIPKDEGWTE"

ORIGIN

```
        1 tgcgcattta tgggcattaa aagggatata aacttaaaat aaggaggaga atatttatgc
       61 cagttacaat aaataatttt aattataatg atcctattga taataataat attattatga
      121 tggagcctcc atttgcgaga ggtacgggga gatattataa agcttttaaa atcacagatc
      181 gtatttggat aataccggaa agatatactt ttggatataa acctgaggat tttaataaaa
      241 gttccggtat ttttaataga gatgtttgtg aatattatga tccagattac ttaaatacta
      301 atgataaaaa gaatatattt ttacaaacaa tgatcaagtt atttaataga atcaaatcaa
      361 aaccattggg tgaaaagtta ttagagatga ttataaatgg tataccttat cttggagata
      421 gacgtgttcc actcgaagag tttaacacaa acattgctag tgtaactgtt aataaattaa
      481 tcagtaatcc aggagaagtg gagcgaaaaa aaggtatttt cgcaaattta ataatatttg
      541 gacctgggcc agttttaaat gaaaatgaga ctatagatat aggtatacaa aatcattttg
      601 catcaaggga aggcttcggg ggtataatgc aaatgaagtt ttgcccagaa tatgtaagcg
      661 tatttaataa tgttcaagaa aacaaaggcg caagtatatt taatagacgt ggatattttt
      721 cagatccagc cttgatatta atgcatgaac ttatacatgt tttacatgga ttatatggca
      781 ttaaagtaga tgatttacca attgtaccaa atgaaaaaaa atttttatg caatctacag
      841 atgctataca ggcagaagaa ctatatacat ttggaggaca agatcccagc atcataactc
      901 cttctacgga taaaagtatc tatgataaag ttttgcaaaa ttttagaggg atagttgata
      961 gacttaacaa ggttttagtt tgcatatcag atcctaacat taatattaat atatataaaa
     1021 ataaatttaa agataaatat aaattcgttg aagattctga gggaaaatat agtatagatg
     1081 tagaaagttt tgataaatta tataaaagct taatgtttgg ttttacagaa actaatatag
     1141 cagaaaatta taaaataaaa actagagctt cttattttag tgattcctta ccaccagtaa
     1201 aaataaaaaa tttattagat aatgaaatct atactataga ggaagggttt aatatatctg
     1261 ataaagatat ggaaaaagaa tatagaggtc agaataaagc tataaataaa caagcttatg
     1321 aagaaattag caaggagcat ttggctgtat ataagataca aatgtgtaaa agtgttaaag
     1381 ctccaggaat atgtattgat gttgataatg aagatttgtt ctttatagct gataaaaata
     1441 gtttttcaga tgatttatct aaaaacgaaa gaatagaata taatacacag agtaattata
     1501 tagaaaatga cttccctata aatgaattaa ttttagatac tgatttaata agtaaaatag
     1561 aattaccaag tgaaaataca gaatcactta ctgattttaa tgtagatgtt ccagtatatg
     1621 aaaaacaacc cgctataaaa aaaattttta cagatgaaaa taccatcttt caatatttat
     1681 actctcagac atttcctcta gatataagag atataagttt aacatcttca tttgatgatg
     1741 cattattatt ttctaacaaa gtttattcat ttttttctat ggattatatt aaaactgcta
     1801 ataaagtggt agaagcagga ttatttgcag gttgggtgaa acagatagta aatgattttg
     1861 taatcgaagc taataaaagc aatactatgg ataaaattgc agatatatct ctaattgttc
     1921 cttatatagg attagcttta aatgtaggaa atgaaacagc taaaggaaat tttgaaaatg
     1981 cttttgagat tgcaggagcc agtattctac tagaatttat accagaactt ttaatacctg
     2041 tagttggagc ctttttatta gaatcatata ttgacaataa aaataaaatt attaaaacaa
     2101 tagataatgc tttaactaaa agaaatgaaa aatggagtga tatgtacgga ttaatagtag
     2161 cgcaatggct ctcaacagtt aatactcaat tttatacaat aaaagaggga atgtataagg
     2221 ctttaaatta tcaagcacaa gcattggaag aaataataaa atacagatat aatatatatt
     2281 ctgaaaaaga aaagtcaaat attaacatcg attttaatga tataaattct aaacttaatg
     2341 agggtattaa ccaagctata gataatataa ataattttat aaatggatgt tctgtatcat
     2401 atttaatgaa aaaaatgatt ccattagctg tagaaaaatt actagacttt gataatactc
     2461 tcaaaaaaaa tttgttaaat tatatagatg aaaataaatt atatttgatt ggaagtgcag
     2521 aatatgaaaa atcaaaagta aataaatact tgaaaaccat tatgccgttt gatctttcaa
     2581 tatataccaa tgatacaata ctaatagaaa tgtttaataa atataatagc gaaatttaa
     2641 ataatattat cttaaattta agatataagg ataataattt aatagattta tcaggatatg
     2701 gggcaaaggt agaggtatat gatggagtcg agcttaatga taaaaatcaa tttaaattaa
     2761 ctagttcagc aaatagtaag attagagtga ctcaaaatca gaatatcata tttaatagtg
```

```
2821 tgttccttga ttttagcgtt agcttttgga taagaatacc taaatataag aatgatggta
2881 tacaaaatta tattcataat gaatatacaa taattaattg tatgaaaaat aattcgggct
2941 ggaaaatatc tattaggggt aataggataa tatggacttt aattgatata aatggaaaaa
3001 ccaaatcggt atttttgaa tataacataa gagaagatat atcagagtat ataaatagat
3061 ggtttttgt aactattact aataatttga ataacgctaa aatttatatt aatggtaagc
3121 tagaatcaaa tacagatatt aaagatataa gagaagttat tgctaatggt gaaataatat
3181 ttaaattaga tggtgatata gatagaacac aatttatttg gatgaaatat ttcagtattt
3241 ttaatacgga attaagtcaa tcaaatattg aagaaagata taaaattcaa tcatatagcg
3301 aatatttaaa agatttttgg ggaaatcctt taatgtacaa taaagaatat tatatgttta
3361 atgcggggaa taaaaattca tatattaaac taaagaaaga ttcacctgta ggtgaaattt
3421 taacacgtag caaatataat caaaattcta aatatataaa ttatagagat ttatatattg
3481 gagaaaaatt tattataaga agaaagtcaa attctcaatc tataaatgat gatatagtta
3541 gaaaagaaga ttatatatat ctagattttt ttaatttaaa tcaagagtgg agagtatata
3601 cctataaata ttttaagaaa gaggaagaaa aattgttttt agctcctata agtgattctg
3661 atgagtttta caatactata caaataaaag aatatgatga acagccaaca tatagttgtc
3721 agttgctttt taaaaaagat gaagaaagta ctgatgagat aggattgatt ggtattcatc
3781 gtttctacga atctggaatt gtatttgaag agtataaaga ttatttttgt ataagtaaat
3841 ggtacttaaa agaggtaaaa aggaaaccat ataatttaaa attgggatgt aattggcagt
3901 ttattcctaa agatgaaggg tggactgaat aatataacta tatgctcagc aaacctattt
3961 tatataagaa aagtttaagt ttataaaatc ttaagtttaa ggatgtagct aaattttgaa
4021 tattagataa actacatgtt t
```

//

APPLIED AND ENVIRONMENTAL MICROBIOLOGY, Aug. 1992, p. 2345–2354
0099-2240/92/082345-10$02.00/0

Vol. 58, No. 8

Molecular Cloning of the *Clostridium botulinum* Structural Gene Encoding the Type B Neurotoxin and Determination of Its Entire Nucleotide Sequence

SARAH M. WHELAN, MICHAEL J. ELMORE, NICOLA J. BODSWORTH, JOHN K. BREHM, TONY ATKINSON, AND NIGEL P. MINTON*

Molecular Genetics Group, Division of Biotechnology, PHLS Centre for Applied Microbiology and Research, Porton Down, Salisbury, Wiltshire SP4 0JG, United Kingdom

Received 11 February 1992/Accepted 5 May 1992

DNA fragments derived from the *Clostridium botulinum* type A neurotoxin (BoNT/A) gene (*botA*) were used in DNA-DNA hybridization reactions to derive a restriction map of the region of the *C. botulinum* type B strain Danish chromosome encoding *botB*. As the one probe encoded part of the BoNT/A heavy (H) chain and the other encoded part of the light (L) chain, the position and orientation of *botB* relative to this map were established. The temperature at which hybridization occurred indicated that a higher degree of DNA homology occurred between the two genes in the H-chain-encoding region. By using the derived restriction map data, a 2.1-kb *BglII-XbaI* fragment encoding the entire BoNT/B L chain and 108 amino acids of the H chain was cloned and characterized by nucleotide sequencing. A contiguous 1.8-kb *XbaI* fragment encoding a further 623 amino acids of the H chain was also cloned. The 3′ end of the gene was obtained by cloning a 1.6-kb fragment amplified from genomic DNA by inverse polymerase chain reaction. Translation of the nucleotide sequence derived from all three clones demonstrated that BoNT/B was composed of 1,291 amino acids. Comparative alignment of its sequence with all currently characterized BoNTs (A, C, D, and E) and tetanus toxin (TeTx) showed that a wide variation in percent homology occurred dependent on which component of the dichain was compared. Thus, the L chain of BoNT/B exhibits the greatest degree of homology (50% identity) with the TeTx L chain, whereas its H chain is most homologous (48% identity) with the BoNT/A H chain. Overall, the six neurotoxins were shown to be composed of highly conserved amino acid domains interceded with amino acid tracts exhibiting little overall similarity. In total, 68 amino acids of an average of 442 are absolutely conserved between L chains and 110 of 845 amino acids are conserved between H chains. Conservation of Trp residues (one in the L chain and nine in the H chain) was particularly striking. The most divergent region corresponds to the extreme carboxy terminus of each toxin, which may reflect differences in specificity of binding to neurone acceptor sites.

Botulinum neurotoxin (BoNT) and tetanus toxin (TeTx) are high-molecular-weight proteins (approximately 150,000 Da) which exert potent neuroparalytic effects on vertebrates (14, 42). They are elaborated by anaerobic gram-positive bacteria belonging to the genus *Clostridium*. TeTx is synthesized by *Clostridium tetani*, whereas the majority of clostridia which produce BoNT are classified as *C. botulinum*. In recent years, however, isolates which resemble *C. barati* and *C. butyricum* have been shown to produce BoNT (15, 24). On the basis of antigenicity, BoNT has been subdivided into seven distinct types, designated A to G. All eight neurotoxins (BoNT/A to BoNT/G and TeTx) are synthesized as a single-chain, 150,000-Da molecule which subsequently becomes nicked to the more potent dichain form, composed of a heavy (H) (approximately 100,000 Da)- and a light (L) (50,000 Da)-chain polypeptide linked by at least one disulfide bridge (40).

It has been proposed (38, 39) that action of both BoNT and TeTx involves three distinct phases. In the first phase the toxins become bound to acceptors on the external surface of the targeted neural cells. This is followed by an energy-dependent internalization step in which the toxin, or part of it, enters the cell. Thereafter, an unidentified active moiety of the toxin causes nerve cell dysfunction by blocking the intracellular release of neurotransmitters. The two classes of toxins differ, however, in that BoNT preferentially inhibits acetylcholine release at the nerve periphery, whereas TeTx blockades the release of inhibitory amino acids principally in the central nervous system. On the basis of a number of pieces of experimental evidence, and by analogy to the characterized binary toxins (e.g., diphtheria and ricin), it is generally assumed that the L chain possesses the catalytic activity responsible for cell poisoning (1, 5, 41) and that the H chain delivers this moiety to the cell cytoplasm by mediating binding of the toxin to the cell and subsequent internalization. The dual role of the H chain in toxicity has been rationalized by the suggestion that the amino-terminal portion mediates internalization (29, 32, 33) and the carboxy terminus plays a crucial role in binding to nerve acceptors (20, 21, 30, 37).

To clarify structural and functional relationships, clostridial neurotoxin gene cloning programs have been initiated in a number of laboratories. As a result of nucleotide sequence analysis of cloned genes, the complete primary sequences of TeTx (11), BoNT/A (4, 44), BoNT/C (17), BoNT/D (3), and BoNT/E (45) are known. In the present report, we describe the cloning of the gene encoding BoNT/B (*botB*) and the derivation of the entire amino acid sequence of the neurotoxin by nucleotide sequencing.

MATERIALS AND METHODS

Bacterial strains, plasmids, and culture conditions. The source of chromosomal DNA was *C. botulinum* Danish, and the recombinant host used for cloning experiments was

* Corresponding author.

Escherichia coli TG1 [Δ(*lac-pro*) *supE thi hsdD5*/F'-*traD36 proA*$^+$ *B*$^+$ *lacI*^{q} *lacZ*ΔM15]. Cloning vectors employed were plasmids pMTL32 (this study), pMTL23 (7), and pCR1000 (26) and the M13 bacteriophages mp18 and mp19 (46). *C. botulinum* was cultivated in USA II broth (2% peptone, 1% yeast extract, 1% N-Z amine, 0.05% sodium mercaptoacetate, 1% glucose [pH 7.4]), and *E. coli* was cultivated in L broth (1% tryptone, 0.5% yeast extract, 0.5% NaCl). Solidified medium (L agar) consisted of L broth with the addition of 2% (wt/vol) Bacto Agar (Difco Laboratories). Antibiotic concentrations used for the maintenance and the selection of transformants were 50 μg of ampicillin (pMTL32) and 50 μg of kanamycin (pCR1000) per ml. Restriction endonucleases and DNA modifying enzymes were purchased from Northumbria Biochemicals Ltd., Taq polymerase was from United States Biochemical Corporation, and radiolabel was from Amersham International.

Purification and manipulation of DNA. Transformation of *E. coli* and large-scale plasmid isolation procedures were as described previously (27). Small-scale plasmid isolation was by the method of Holmes and Quigley (19), while chromosomal DNA from *C. botulinum* was prepared essentially as described by Marmur (23). Restriction endonucleases and DNA modifying enzymes were used under the conditions recommended by the suppliers. Digests were electrophoresed in 1% agarose slab gels on a standard horizontal system (Bethesda Research Laboratories model H4), employing Tris borate-EDTA (0.09 M Tris borate, 0.002 M EDTA) buffer. Fragments were isolated from gels by electroelution (25). All primary cloning procedures were undertaken under United Kingdom ACGM C2 containment conditions, and total cell lysates of all recombinants carrying cloned material were tested in mice for the absence of toxic polypeptides.

DNA-DNA hybridization experiments. DNA restriction fragments were transferred from agarose gels to Zeta Probe nylon membrane by the procedure of Reed and Mann (34). After partial depurination with 0.25 M HCl (15 min), DNA was transferred in 0.4 M NaOH by capillary elution for 4 to 16 h. Bacterial colonies were screened for desired recombinant plasmids by in situ colony hybridization (13), using nitrocellulose filter disks (0.22 μm; Schleicher and Schuell). The gel-purified *botA* DNA fragments were labelled with [α-^{32}P]dATP, using a multiprime kit supplied by Amersham International. Hybridizations were carried out as described previously (44) at temperatures ranging from 45 to 60°C.

Nucleotide sequence of pCBB plasmid inserts. The insert of plasmid pCBB1 was excised by cleavage with *Bam*HI and *Bgl*II and circularized by treatment with T4 ligase, and size-fractionated 500- to 1,000-bp fragments generated by sonication were cloned into the *Sma*I site of M13mp18 (for experimental conditions, see reference 28). Approximately 50 templates were then sequenced by the dideoxynucleotide method of Sanger et al. (35), using a modified version of bacteriophage T7 DNA polymerase, Sequenase (43). Experimental conditions used were those stated by the supplier (United States Biochemical Corp.). The inserts of plasmids pCBB2 and pCBB3 were sequenced by using templates derived by subcloning the entire region between the appropriate sites of M13mp18 and M13mp19. Sequence data obtained by employing universal primer were then sequentially extended by the use of custom-synthesized oligonucleotide primers. In certain instances, templates were generated by the insertion of *Dra*I restriction subfragments into the *Sma*I site of M13mp18. In all cases the sequence was determined on both DNA strands. The chromosomal DNA region amplified with primers X1 and X2 (Fig. 1) was cloned directly into ddT-tailed, *Sma*I-cut M13mp8 (prepared by incubating *Sma*I-cut DNA with terminal transferase in the presence of dideoxy TTP), and the resultant template was sequenced with universal primer. DNA sequence data were analyzed by using the computer software of DNASTAR Inc.

Amplification of DNA by PCR. Amplification of *C. botulinum* DNA was undertaken by polymerase chain reaction (PCR), using an M J Research Inc. thermal cycler. Reaction mixtures contained 10 mM Tris-HCl, 50 mM KCl, 3 mM $MgCl_2$, 0.1 mM deoxynucleoside triphosphate, 30 nmol of each primer, 2.5 U of *Taq* polymerase, and 10 ng of strain Danish genomic DNA, in a final volume of 0.1 ml. Amplification was for 30 cycles, as follows: 1.5 min at 93°C, 3 min at 37°C, and 3 min at 72°C. For inverse PCR, 140 ng of chromosomal DNA, cleaved with an appropriate restriction endonuclease, was ligated overnight at 14°C in a 50-μl volume and a 10-μl portion of the resultant concatenated DNA was used in PCR.

Nucleotide sequence accession number. The nucleotide sequence has been submitted to the GenBank/EMBL data banks, with the accession number M81186.

RESULTS AND DISCUSSION

Southern blot analysis of the *botB* gene. Previous studies have shown that BoNT appears to conform to the classical A-B binary toxin model (12). Thus, both L and H chains are required for toxicity (14, 39). The risk of generating an *E. coli* clone with the capacity to produce a neuroparalytic polypeptide may therefore be alleviated by cloning genomic restriction fragments which encode principally only one component of the dichain molecule. To identify such fragments, we exploited DNA homology between *botB* and the previously cloned *botA* (44).

A 389-bp *Hpa*I-*Xho*II *botA* fragment, encoding amino acids 216 through 346 of the BoNT/A L chain, and a 628-bp *Hae*II-*Hin*dIII fragment, coding for amino acids 526 through 736 of the H chain (44), were radiolabelled and used in DNA-DNA hybridizations with type B chromosomal DNA cleaved with various restriction enzymes. Reactions were performed in aqueous solution over a range of temperatures. "Weak" hybridization between the two genes was found to occur at 53 and 56°C with the L- and H chain probes, respectively (data not shown). The strength of the signal observed and the relatively low stringency required were indicative of a fairly low level of DNA homology between *botA* and *botB*. Furthermore, these results suggest that the L-chain-encoding regions of the two genes are less homologous than the H-chain-encoding region, at least in the areas probed. The conditions under which hybridization occurred having been established, the type B genomic DNA was cleaved with various combinations of restriction endonucleases and the nylon membranes carrying the resultant fragments were sequentially hybridized with the two probes. The data obtained allowed the derivation of a restriction map of the region of the type B genome encoding *botB*. Furthermore, the use of the two probes enabled the assignment of both the position of *botB* and its relative orientation with respect to the derived map (Fig. 1).

Cloning and sequencing of the *botB* L chain. The restriction map derived by the Southern blot experiments (Fig. 1)

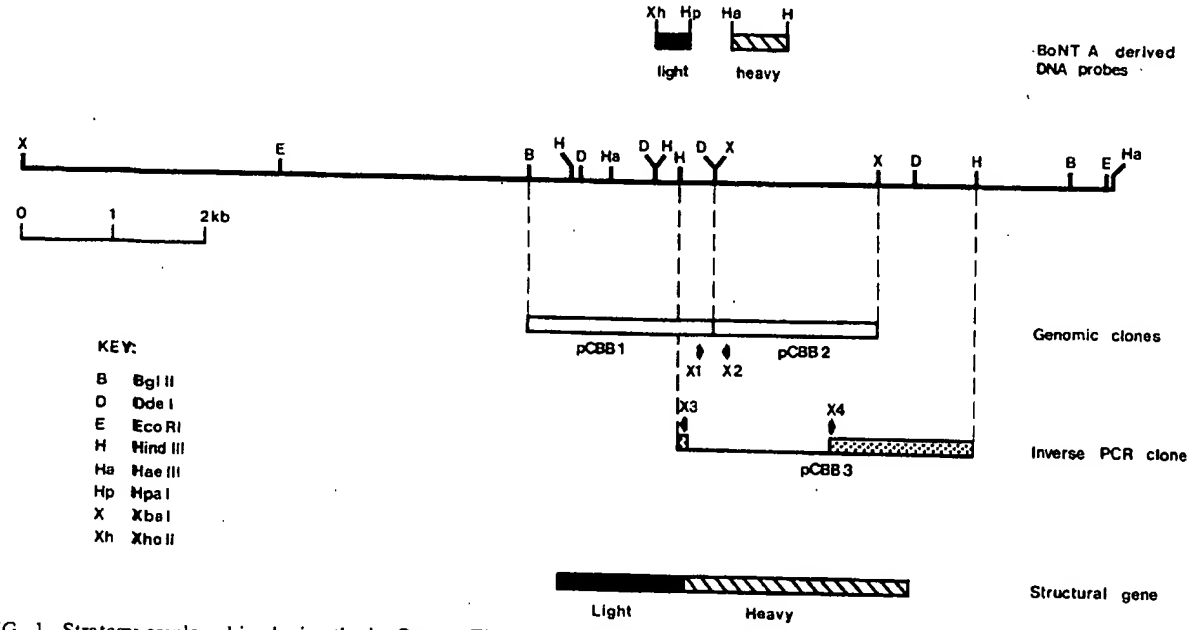


FIG. 1. Strategy employed in cloning the *botB* gene. The illustrated restriction map of the *C. botulinum* genome was generated by using the indicated *botA* DNA fragments as probes in Southern blots. Regions of the strain B/Danish chromosome that were cloned in recombinant plasmids pCBB1 and pCBB2 are represented by open boxes below the restriction map. The cloned inserts of these plasmids were shown to be contiguous on the genome by PCR amplification of the region of the chromosome spanning their common *Xba*I site, using primers X1 (5'-CCAAGTGAAAATACAGAATCAC-3') and X2 (3'-CCCACTTTGTCTATCATTTA-5') and sequencing across this junction. The insert of pCBB3 was derived by PCR amplification of *Hin*dIII-cut, concatenated chromosomal DNA, using primers X4 (5'-ATAGAGATTTATATAT TGGAG-3') and X3 (5'-TTATATACAGCCAAATGCTCCTTGC-3').

indicated that a 2.1-kb *Bgl*II-*Xba*I fragment principally encoded the L chain of BoNT/B. To clone this DNA, and to minimize the risk of cloning contiguous BoNT/B-encoding regions, the targeted fragment was purified by a two-stage gel isolation procedure. *C. botulinum* type B chromosomal DNA was cleaved with *Xba*I, and fragments of approximately 7.5 kb were purified from agarose gels by electroelution. The isolated DNA was then subjected to digestion with *Bgl*II, DNA fragments of around 2.1 kb were gel purified and ligated to pMTL32 vector DNA (Fig. 2) cut with *Xba*I and *Bam*HI, and the resultant TG1 transformants were screened for the presence of recombinant clones, using the *botA* L-chain probe. Vector pMTL32 was specifically constructed for the purposes of cloning the *botB* DNA (Fig. 2). Based on the pMTL1003 backbone (6), it carries multiple cloning sites flanked on either side by tandem copies of transcriptional terminators. Heterologous genes inserted into the multiple cloning sites will therefore only be expressed if they carry indigenous transcriptional elements recognized by the RNA polymerases of *E. coli*.

The recombinant plasmid obtained, designated pCBB1, was shown by digestion with appropriate endonucleases to contain restriction enzyme recognition sites consistent with the map illustrated in Fig. 1. Its entire insert was excised by digestion with *Bam*HI and *Bgl*II, and M13 recombinant templates containing random inserts were derived by using a sonication procedure (28). By using these templates and custom synthesized oligonucleotides, the entire nucleotide sequence of the insert was determined on both strands. Translation of the resultant sequence indicated the presence of an open reading frame encoding a polypeptide of 549 amino acids in size. The amino terminus of this polypeptide exhibited perfect conformity to that experimentally determined for purified BoNT/B L chain (36). Amino acids 442 through 459 were identical to those determined for purified BoNT/B H chain (36). Thus, the insert carried by pCBB1 was deemed to encode the entire L chain of BoNT/B and 108 amino acids from the H chain.

Cloning and sequencing of the *botB* H chain. After it was determined that the 2.1-kb *Bgl*II-*Xba*I fragment encoded the entire BoNT/B L chain and the amino terminus of the H chain, it was apparent that the adjacent 1.8-kb *Xba*I fragment (Fig. 1) should encode the majority of the remaining H chain. Type B chromosomal DNA was cleaved with *Hin*dIII, fragments of approximately 3.5 kb were isolated and digested with *Xba*I, and fragments of around 1.8 kb were gel purified. The isolated DNA was ligated with *Xba*I-cleaved pMTL32 and transformed into *E. coli* TG1, and recombinant plasmids were identified by probing with the radiolabelled *botA* H-chain probe. One such plasmid was designated pCBB2, and the nucleotide sequence of its insert was determined, following its insertion into M13mp18, by employing custom-synthesized oligonucleotide primers.

Translation of the nucleotide sequence obtained revealed the presence of a continuous open reading frame of 623 codons, in the same reading frame relative to the *Xba*I site of that of the insert of plasmid pCBB1. To confirm that the two sequences were indeed contiguous, a 289-bp region of DNA encompassing the *Xba*I site was amplified from type B genomic DNA by using primers X1 and X2 (Fig. 1) in a PCR and cloned directly into ddT-tailed *Sma*I-cut M13mp8. Nu-

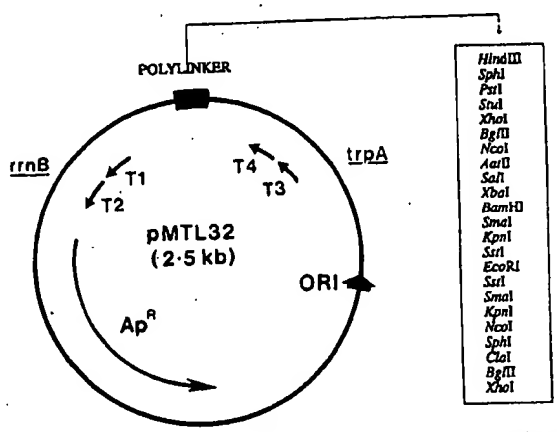


FIG. 2. Cloning vector pMTL32. This plasmid was derived as follows. A synthetic DNA fragment (5'-AGCCCGCCTAATGAGCG GGCTTTTTTT-3'), corresponding to the *E. coli trpA* transcriptional terminator, was ligated to *Stu*I-cleaved pMTL23 (7), and a recombinant plasmid (pTRP23) was selected in which two tandem copies of *trpA* had been inserted. The resultant double terminator, together with part of the pMTL23 polylinker region, was excised as a 107-bp *Nru*I-*Eco*RI fragment and inserted between the *Eco*RI and *Eco*RV sites of plasmid pMTL1003 (6). As the ca. 350-bp *Eco*RI-*Eco*RV fragment of pMTL1003 is deleted during this manipulation, the resultant plasmid, pMTL32, does not carry a copy of the *trp* promoter.

cleotide sequencing of a derivative template, using universal primer, demonstrated that the inserts of plasmids pCBB1 and pCBB2 were contiguous in the *C. botulinum* type B chromosome.

Completion of the *botB* sequence. By combining the two sequences of pCBB1 and pCBB2, the derived contiguous open reading frame encoded 1,170 amino acids, indicating that some 120 or so codons of the *botB* gene were missing. A DNA region encompassing the remaining 3' end of the gene was cloned by inverse PCR. Type B chromosomal DNA was cleaved with *Hin*dIII and incubated with T4 ligase, and the resultant concatenated DNA was used as a template in PCR with oligonucleotides X3 and X4 (Fig. 1). The 1.6-kb fragment generated was cloned directly into the specialized vector pCR1000, and the recombinant plasmid obtained was designated pCBB3. A plasmid sequence reaction, undertaken with a primer previously employed in the determination of the nucleotide sequence of the insert of plasmid pCBB2, confirmed the presence of the *botB* gene. Thereafter, the nucleotide sequence of the region of pCBB3 encompassing the 3' end of *botB* was determined by subcloning selected overlapping fragments into M13. To rule out the possibility that the insert of pCBB3 may have contained PCR-induced errors, a second version of this plasmid recombinant was derived by cloning the amplified DNA product from a further independent inverse PCR. Nucleotide sequencing of the appropriate regions of this second plasmid gave a sequence identical to that already derived from the primary isolate of pCBB3.

The entire nucleotide sequence of the *botB* gene (Fig. 3) was obtained by splicing the individual sequence information derived from the inserts of pCBB1, pCBB2, and pCBB3 into a contiguous sequence. The gene is composed of 1,291 codons, initiating with an AUG codon at position 55 and terminating with a UAA stop codon at position 3928 (Fig. 3). The choice of these particular translational codons is typical of clostridial genes (47). As with all other *bot* genes characterized to date, the high A+T content of the DNA (74.6%) results in an extreme bias towards the use of codons ending in A or T and the frequent use of codons recognized as modulators in *E. coli*. The translational start codon is preceded by a sequence typical of clostridial ribosome binding sites (47).

Alignment of the nucleotide sequences of the two *botA*-derived DNA probes used in Southern blot mapping with the equivalent regions of *botB* confirmed that the greater degree of homology existed in the respective H-chain-encoding regions over those encoding L chain. Specifically, the 628-bp *Hae*III-*Hin*dIII *botA* fragment demonstrated 65% homology with *botB*, whereas the 389-bp *Hpa*I-*Xho*II *botA* fragment had 54.8% homology with *botB*. Comparative alignment demonstrated that, in general, the overall DNA homology (Table 1) between the H- and L-chain-encoding regions of all sequenced neurotoxin genes reflected the level of amino acid sequence homology (Table 2) and averaged between 50 and 60% identity. One consequence of this relative dissimilarity between genes is that DNA probes specific to each toxin gene may be easily designed. However, although there is sufficient homology in certain regions to derive a generalized probe for the generic detection of neurotoxin genes, it has not proven possible to design a probe which hybridizes to all *bot* genes and not to the TeTx gene (unpublished data).

Predicted amino acid sequence of BoNT/B. The deduced primary sequence of BoNT/B demonstrates that the toxin is composed of 1,291 amino acid residues. By comparison to partial amino acid sequences derived from purified polypeptides from other *C. botulinum* type B strains, it is apparent that variations in toxin structure occur. Thus, although amino acid residues 2 through 17 exhibit perfect conformity to the sequence derived by Edman degradation of purified BoNT/B L chain of strain B/Okra (36), the amino acid at position 23 of the H chain was determined (10) to be Arg rather than the Ser residue seen here (position 464, Fig. 4). Similarly, the BoNT/B of strain B/657 possesses a Met amino acid at position 30 of the L chain (9) compared with Thr in the case of BoNT/B from both strain Danish and strain B/Okra. Variations in the primary amino acid sequence of other types of BoNT have been noted, e.g., between BoNT/A of strains 62A (4) and NCTC 2916 (44) and between BoNT/E of strains Beluga, Mashike, Iwanai, Otaru, and NCTC 11219 (see reference 45). In the case of BoNT/B, such variations help to explain observed dissimilarity in the immunological properties of BoNT/B isolated from different strains (16, 31).

Pairwise comparisons of the respective L- and H-chain components of all six toxins were undertaken, and the results are summarized in Table 2. From this it can be seen that, with notable exceptions, the overall level of identity between L chains varies from around 30 to 35%. The three exceptions are the degrees of homology seen between BoNT/E and TeTx (40%), BoNT/C and BoNT/D (47%), and BoNT/B and TeTx (50%). The last homology is particularly striking and serves to illustrate the close relationships between the pharmacological action of BoNT and TeTx. In contrast to the situation with the L-chain subunit, the H chains of BoNT/B and TeTx represent one of the most divergent pairings. The greatest level of homology (48% identity) to BoNT/B in this region is with BoNT/A. A similar relationship exists between the dichain components of BoNT/E and TeTx and BoNT/A. These observations sug-

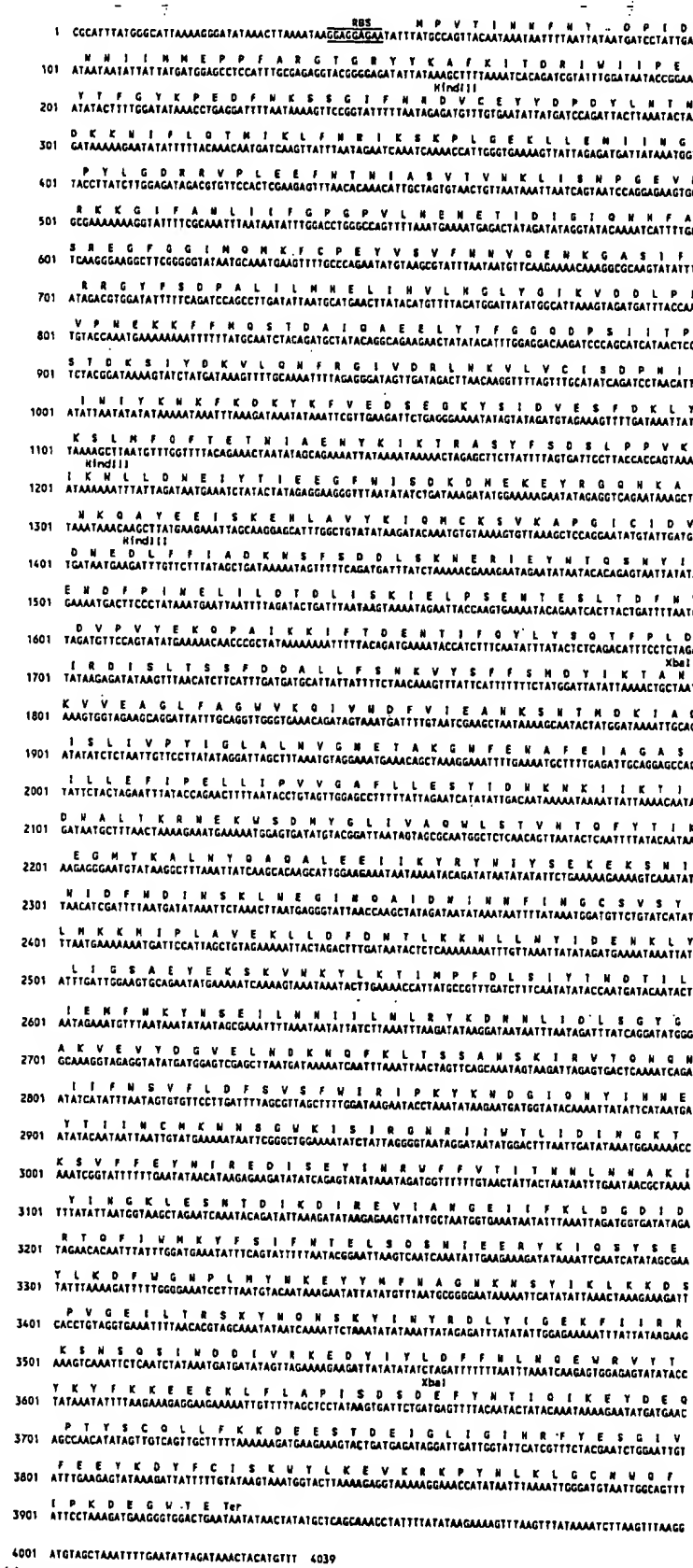
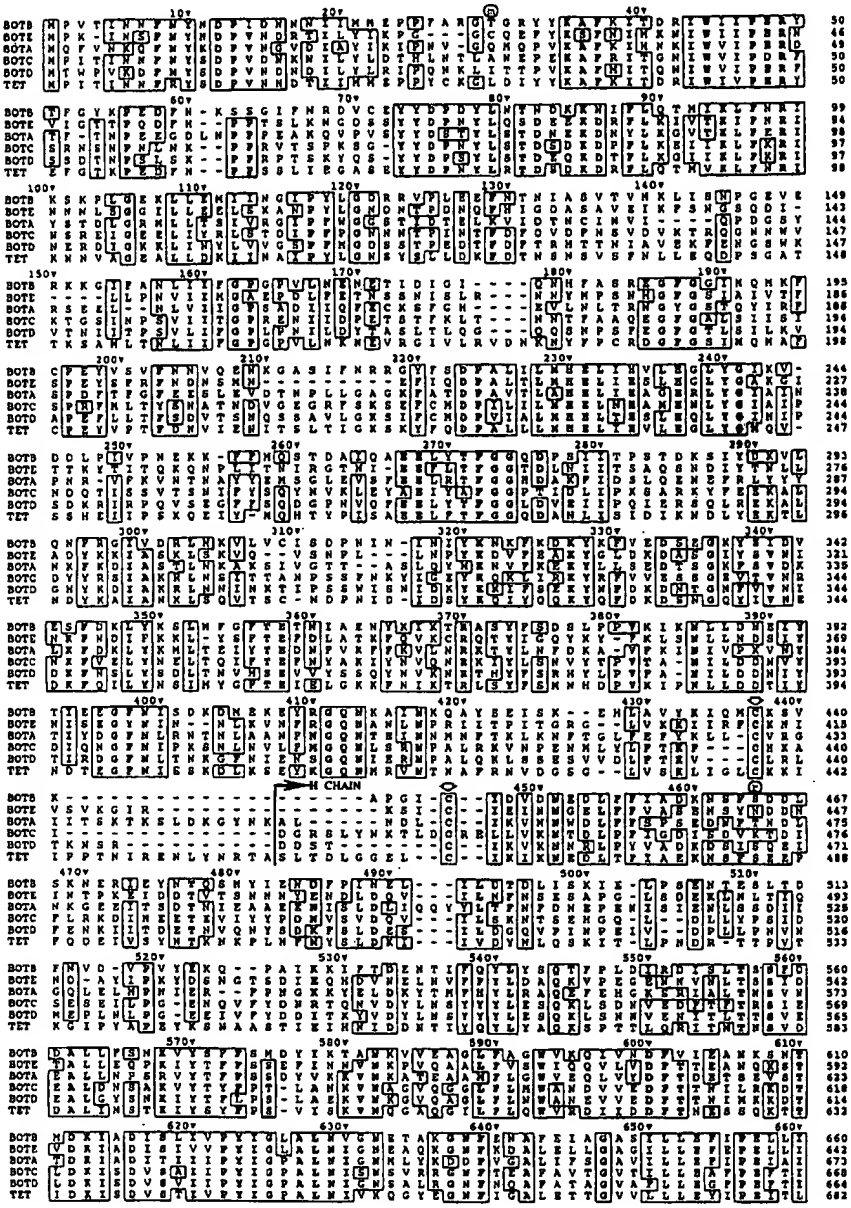


FIG. 3. Complete nucleotide sequence of *botB*. The illustrated sequence was derived by amalgamation of the derived nucleotide sequences of the inserts of pCBB1, pCBB2, and pCBB3 (Fig. 1). The BoNT/B amino acid sequence is given in the single-letter code above the first nucleotide of the corresponding codon. The ribosome binding site is indicated by a line above and below the sequence.



gest that either L- and H-chain domains of an individual neurotoxin have evolved at disproportionate rates or at various stages toxins have arisen by fusion of distinct H and L chains.

A full alignment of all presently characterized clostridial neurotoxins is illustrated in Fig. 4. This analysis demonstrates that they are composed of highly conserved domains interceded with tracts of amino acids exhibiting little overall relatedness, although considerable identity between the components of a specific pair is apparent in certain of these regions. The close relationship that exists between these neurotoxins is further exemplified by their arrangement of polar and nonpolar amino acids. Thus, analysis by the method of Kyte and Doolittle (22) demonstrates that all six neurotoxins possess a highly characteristic hydrophilicity profile (Fig. 5). From these profiles it is apparent that a substantial region of hydrophobicity is centrally conserved in all toxins (Fig. 5). This region has previously been suggested, in the case of TeTx and BoNT/A (11, 44), to play some role in the channel-forming properties of toxin H chain.

Within the L-chain region (average size, 442 amino acids), 68 amino acids are totally conserved, including a Cys amino acid in the carboxy termini. The Cys residues at this position represent one of only two positions where this particular amino acid is absolutely conserved, the second occurring in the amino termini of the H chains. These two Cys residues are therefore undoubtedly involved in the formation of the disulfide bridge linking the two dichain components of all six neurotoxins. Eleven of the conserved amino acids of the

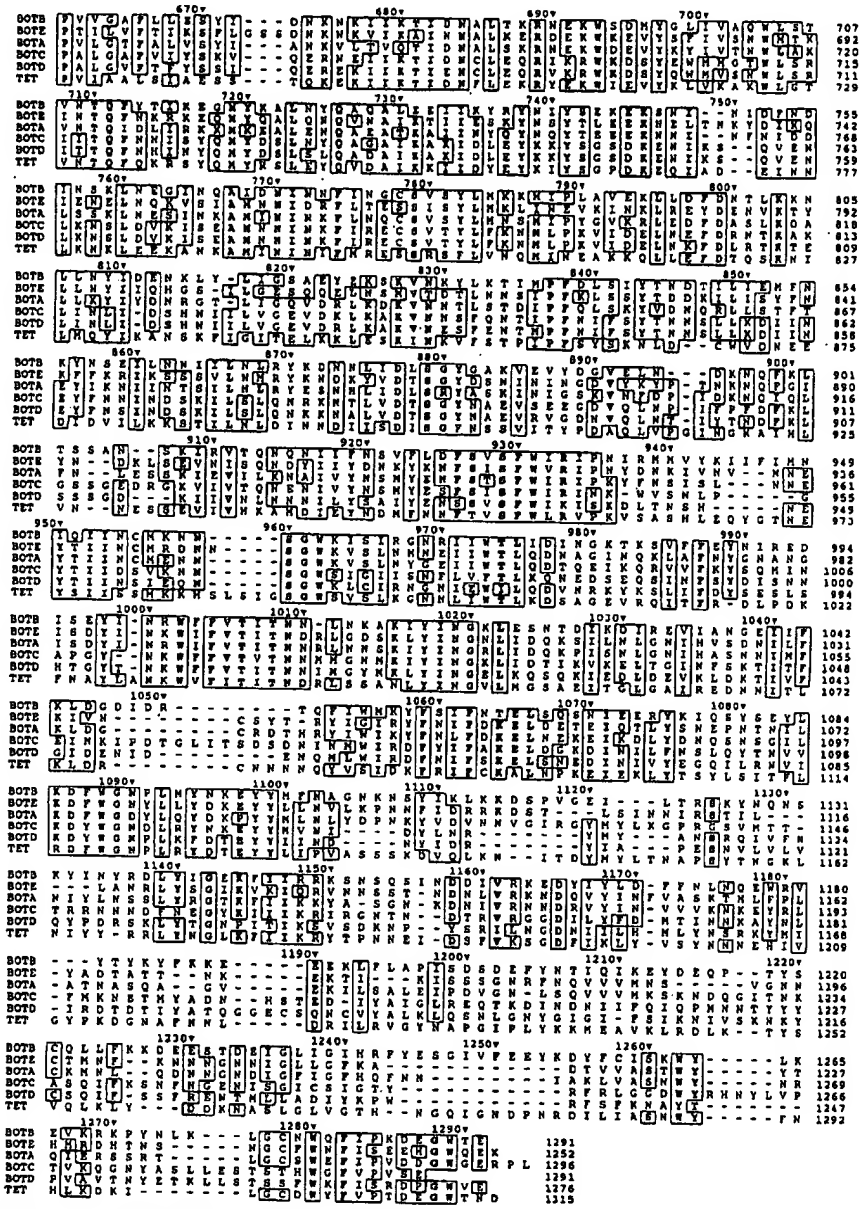


FIG. 4. Comparative alignment of clostridial neurotoxins. The illustrated alignment was essentially derived by using computer program CLUSTAL (18) and has been gapped to maximize similarity. Regions highly conserved among all six neurotoxins have been boxed and include areas in which conservative replacements have occurred, in addition to sequence identity. Amino acids absolutely conserved in five of six toxins are in boldface type. Numbering above the alignment corresponds to that of BoNT/B. Differences from the partial amino acid sequence of BoNT/B of strain B/657 (Met-30 and Arg-464) are circled and indicated above the strain B/Danish BoNT/B sequence. The Cys amino acids presumed to be involved in the formation of the disulfide bridge between the neurotoxin L and H chains are marked by downward facing open arrows.

neurotoxin L chains reside in a region (positions 223 to 241 of BoNT/B) which encompasses a histidine-rich motif. The three conserved His residues of this region, on the basis of their conservation in BoNT/A, BoNT/E, and TeTx, have previously been suggested to play some role in the presumed catalytic activity of the L chain (4). Their conservation in all six neurotoxins does not detract from this hypothesis. Preliminary work, however, in which site-directed mutagenesis has been used to effect amino acid substitutions at all three His positions did not affect the toxicity of a BoNT/A subunit in an *Aplysia californica* buccal ganglian model system (2).

A total of 110 amino acids are absolutely conserved within the H-chain region (average size, 845 amino acids). Most notable is the high degree of conservation of Trp amino acids. Of the 13 Trp residues which occur in the BoNT/B H

TABLE 1. Nucleotide sequence homology between characterized *bot* structural genes[a]

Gene	% Identity among H-chain- and among L-chain-encoding regions[b]					
	botA	*botB*	*botC*	*botD*	*botE*	TET
botA		58.6	52.8	55.2	62.7	54.0
botB	50.1		54.1	55.3	58.4	56.3
botC	49.8	52.5		70.0	53.0	48.5
botD	50.8	52.1	61.5		52.5	53.8
botE	51.6	51.2	51.2	52.3		53.3
TET	49.3	65.2	51.1	53.5	49.8	

[a] A, B, C, D, and E refer to the respective gene; TET represents the TeTx gene.
[b] Identities between H-chain-encoding and between L-chain-encoding regions are given above and below the diagonal, respectively.

TABLE 2. Degree of homology between the respective L- and H-chain components of characterized clostridial neurotoxins[a]

Neurotoxin	% Identity among H chains and among L chains[b]					
	A	B	C	D	E	TET
A		48	34	35	46	35
B	31		39	40	44	36
C	32	32		56	36	32
D	35	35	47		37	36
E	33	33	32	33		35
TET	30	50	34	34.5	40	

[a] A, B, C, D, and E refer to the respective BoNT; TET represents TeTx.
[b] Identities between H chains and between L chains are given above and below the diagonal, respectively.

chain, 9 are absolutely conserved in all toxins. In the majority of the four other positions, where a difference does exist in a particular toxin in six of nine cases, the substitution of a chemically similar amino acid has occurred. The only Trp that occurs in the BoNT/B L chain is conserved in all neurotoxins. The functional significance of the apparent evolutionary pressure for maintaining this relatively rare amino acid, or chemically similar residues, at these positions in BoNT and TeTx remains unknown. However, previous studies in which BoNT Trp residues have been selectively modified by chemical means has established a potential role in both toxicity and immunogenicity (8). Indeed, in one study it was reported that the modification of a single Trp resulted in nearly complete detoxification (Shibaeva et al., 1981, cited in reference 8). The selective disruption of conserved Trp amino acids in BoNT by site-directed mutagenesis should clarify which residue(s), if any, is important in toxicity and antigenicity.

The most striking area of sequence divergence between toxins occurs in carboxy-terminal areas of their H chains from, in the case of BoNT/B, around residue 1100 onwards. Given that this part of the toxin plays a major role in cell binding and that different toxins bind to distinct cell acceptor molecules, the finding that none of the toxins are alike in this region is perhaps not surprising. In view of the preceding region of divergence, the conservation of a sequence motif conforming to the consensus W-X-F-I/V-P/S-X-D/E-X-G-W-X-E/N (BoNT/B positions 1280 through 1291) at the extreme carboxy terminus is particularly intriguing.

ACKNOWLEDGMENTS

We thank Helen King and Howard Tranter for providing cell paste, Cliff Shone and Jack Melling for assisting with toxicity testing

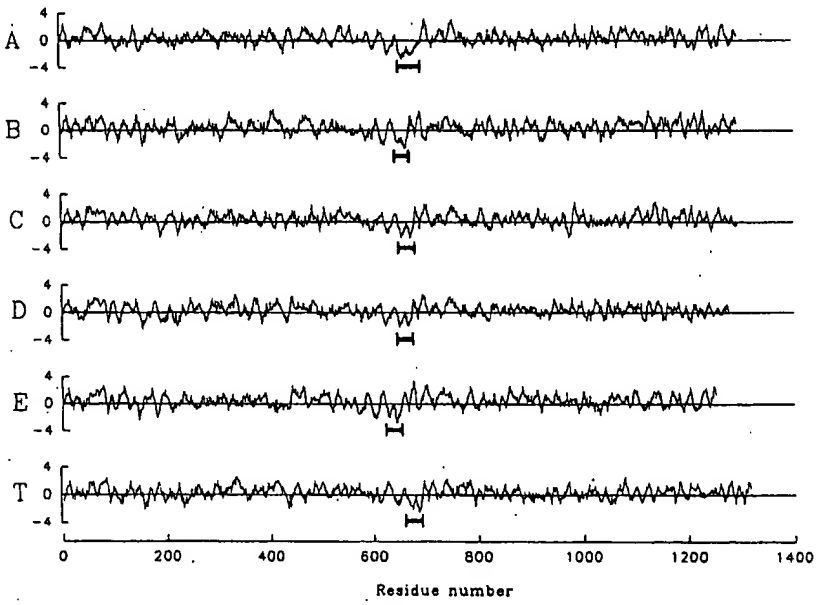


FIG. 5. Hydrophobicity plots of all currently characterized clostridial neurotoxins. Hydrophobicity was calculated by using the computer program of Kyte and Doolittle (22) with a window size of nine amino acids. The average value for each toxin was as follows: BoNT/A, −0.37; BoNT/B, −0.42; BoNT/C, −0.41; BoNT/D, −0.36; BoNT/E, −0.45; TeTx., −0.37. The conserved hydrophobic region is indicated below each profile by a barred line. The respective residues involved are 652 through 687 (BoNT/A), 642 through 671 (BoNT/B), 648 through 678 (BoNT/C), 646 through 674 (BoNT/D), 624 through 654 (BoNT/E), and 660 through 691 (TeTx).

of recombinant clones, and Nicola Minion for typing the manuscript.

This work was supported by the Public Health Laboratory Service, the U.K. Ministry of Defence (grant 2238/002), and the U.S. Army Medical Research and Development Command (grant DAMD17-90-Z-0033).

REFERENCES

1. **Ahnert-Higler, G., U. Weller, M. E. Dauzenroth, E. Habermann, and M. Gratzl.** 1989. The tetanus toxin light chain inhibits exocytosis. FEBS Microbiol. Lett. **242**:245–248.
2. **Binz, T., O. Grebenstein, H. Kurazono, J. Thierer, U. Eisel, K. Wernars, M. Popoff, S. Mochida, B. Poulain, L. Tauc, S. Kozaki, and H. Niemann.** 1992. Molecular biology of the L chains of clostridial neurotoxins, p. 56–65. *In* B. Witholt, J. E. Alouf, G. J. Boulnois, P. Cossart, B. W. Dijkstra, P. Falmagne, F. J. Fehrenbach, J. Freer, H. Niemann, R. Rappuoli, and T. Wadström (ed.), Bacterial protein toxins. Gustav Fischer Verlag, New York.
3. **Binz, T., H. Kurazono, M. R. Popoff, M. W. Eklund, G. Sakaguchi, S. Kozaki, K. Krieglstein, A. Henschen, D. M. Gill, and H. Niemann.** 1990. Nucleotide sequence of the gene encoding *Clostridium botulinum* neurotoxin type D. Nucleic Acids Res. **18**:5556.
4. **Binz, T., H. Kurazono, M. Wille, J. Frevert, K. Wernars, and H. Niemann.** 1990. The complete sequence of botulinum neurotoxin type A and comparison with other clostridial neurotoxins. J. Biol. Chem. **265**:9153–9158.
5. **Bittner, M. A., B. DasGupta, and R. W. Holz.** 1989. Isolated light chains of botulinum neurotoxins inhibit exocytosis. J. Biol. Chem. **264**:10354–10360.
6. **Brehm, J. K., S. P. Chambers, K. J. Bown, T. Atkinson, and N. P. Minton.** 1991. Molecular cloning and nucleotide sequence determination of the *Bacillus stearothermophilus* NCA 1503 superoxide dismutase genes and its overexpression in *Escherichia coli*. Appl. Microbiol. Biotechnol. **36**:358–363.
7. **Chambers, S. P., S. E. Prior, D. A. Barstow, and N. P. Minton.** 1988. The pMTL *nic*$^-$ cloning vectors. I. Improved pUC polylinker regions to facilitate the use of sonicated DNA for nucleotide sequencing. Gene **68**:139–149.
8. **DasGupta, B. R.** 1990. Structure and biological activity of botulinum neurotoxin. J. Physiol. Paris. **84**:220–228.
9. **DasGupta, B. R., and A. Datta.** 1987. Botulinum neurotoxin type B (strain 657); partial sequence and similarity with tetanus toxin. Fed. Proc. Fed. Am. Soc. Exp. Biol. **45**:2289.
10. **DasGupta, B. R., and A. Datta.** 1988. Botulinum neurotoxin type B (strain 657), partial sequence and similarity to tetanus neurotoxin. Biochimie **70**:811–817.
11. **Eisel, U., W. Jarausch, K. Goretzki, A. Henschen, J. Engels, U. Weller, M. Hudel, E. Habermann, and H. Niemann.** 1986. Tetanus toxin: primary structure, expression in *E. coli* and homology with botulinum toxins. EMBO J. **5**:2495–2502.
12. **Gill, D. M.** 1978. Seven toxic peptides that cross cell membranes, p. 291–332. *In* J. Jeljaszewicz and T. Wadstrom (ed.), Bacterial toxins and cell membranes. Academic Press, Inc., New York.
13. **Grunstein, M., and D. Hogness.** 1975. Colony hybridisation: a simple method for the isolation of cloned DNAs that contain a specific gene. Proc. Natl. Acad. Sci. USA **72**:3961–3965.
14. **Habermann, E., and F. Dreyer.** 1986. Clostridial neurotoxins: handling and action at the cellular and molecular level. Curr. Top. Microbiol. Immunol. **129**:93–179.
15. **Hall, J. D., L. McCroskey, B. J. Pincomb, and C. L. Hatheway.** 1985. Isolation of an organism resembling *Clostridium barati* which produces type F botulinum toxin from an infant with botulism. J. Clin. Microbiol. **21**:654–655.
16. **Hatheway, C. L., L. M. McCroskey, G. L. Lombard, and V. R. Dowell, Jr.** 1981. Atypical toxin variant of *Clostridium botulinum* type B associated with infant botulism. J. Clin. Microbiol. **14**:607–611.
17. **Hauser, D., M. W. Eklund, H. Kurazono, T. Binz, H. Niemann, D. M. Gill, P. Boquet, and M. R. Popoff.** 1990. Nucleotide sequence of the *Clostridium botulinum* C1 neurotoxin Nucleic Acids Res. **18**:4924.
18. **Higgins, D. G., and P. M. Sharp.** 1988. CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. Gene **73**:237–244.
19. **Holmes, D. S., and M. Quigley.** 1981. A rapid boiling method for the preparation of bacterial plasmids Anal. Biochem. **114**:193–197.
20. **Kozaki, S., A. Miki, Y. Kamata, J. Ogasawara, and G. Sakaguchi.** 1989. Immunological characterization of papain-induced fragments of *Clostridium botulinum* type A neurotoxin and interaction of the fragments with brain synaptosomes. Infect. Immun. **57**:2634–2639.
21. **Kozaki, S., J. Ogasawara, Y. Shimote, Y. Kamata, and G. Sakaguchi.** 1987. Antigenic structure of *Clostridium botulinum* type B neurotoxin and its interaction with gangliosides, cereboside, and free fatty acids. Infect. Immun. **55**:3051–3056.
22. **Kyte, J., and R. F. Doolittle.** 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157**:105–132.
23. **Marmur, J.** 1961. A procedure for the isolation of deoxyribonucleic acid from microorganisms. J. Mol. Biol. **3**:208–218.
24. **McCroskey, L. M., C. L. Hatheway, L. Fenicia, B. Pasolini, and P. Aureli.** 1985. Characterization of an organism that produces type E botulinum toxin but which resembles *Clostridium butyricum* from the feces of an infant with type E botulism. J. Clin. Microbiol. **23**:210–202.
25. **McDonnell, M. W., M. N. Simon, and F. W. Studier.** 1977. Analysis of restriction fragments of T7 DNA and determination of molecular weights by electrophoresis in neutral and alkaline gels. J. Mol. Biol. **110**:119–146.
26. **Mead, D. A., N. C. Pey, C. Herrnstadt, R. A. Marcil, and L. M. Smith.** 1991. A universal method for the direct cloning of PCR amplified nucleic acid. Bio/Technology **9**:657–663.
27. **Minton, N. P., T. Atkinson, and R. F. Sherwood.** 1983. Molecular cloning of the *Pseudomonas* carboxypeptidase G_2 gene and its expression in *Escherichia coli* and *Pseudomonas putida*. J. Bacteriol. **156**:1222–1227.
28. **Minton, N. P., H. M. S. Bullman, M. D. Scawen, T. Atkinson, and H. J. Gilbert.** 1986. Nucleotide sequence of the *Erwinia chrysanthemi* 1066 L-asparaginase gene. Gene **46**:25–35.
29. **Mochida, S., B. Poulain, U. Weller, E. Habermann, and L. Tauc.** 1989. Light chain of tetanus toxin intracellularly inhibits acetylcholine release at neuro-neuronal synapses, and its internalisation is mediated by heavy chain. FEBS Microbiol. Lett. **253**: 47–51.
30. **Morris, N. P., E. Consiglio, L. D. Kohn, W. H. Habig, M. C. Hardegree, and T. B. Helting.** 1981. Interaction of fragments B and C of tetanus toxin with neural and thyroid membranes and with gangliosides. J. Biol. Chem. **255**:6071–6076.
31. **Notermans, S., S. Kozaki, Y. Kamata, and G. Sakaguchi.** 1984. Use of monoclonal antibodies in enzyme linked immunosorbent assay (ELISA) for detection of botulinum type B toxins. Jpn. J. Med. Sci. Biol. **37**:138–140.
32. **Poulain, B., S. Mochida, J. D. F. Wadsworth, U. Weller, B. Hogy, E. Habermann, J. O. Dolly, and L. Tauc.** 1990. Inhibition of neurotransmitter release by botulinum neurotoxins and tetanus toxin at *Aplysia* synapses: role of the constituent chains. J. Physiol. (Paris) **84**:247–261.
33. **Poulain, B., S. Mochida, U. Weller, B. Hogy, E. Habermann, J. D. F. Wadsworth, C. C. Shone, J. O. Dolly, and L. Tauc.** 1991. Heterologous combinations of heavy and light chains from botulinum neurotoxin A and tetanus toxin inhibit neurotransmitter release in *Aplysia*. J. Biol. Chem. **266**:9580–9585.
34. **Reed, K. C., and D. A. Mann.** 1985. Rapid transfer of DNA from agarose to nylon membranes. Nucleic Acids Res. **13**:7207–7221.
35. **Sanger, F., A. R. Coulson, B. G. Barrell, A. J. H. Smith, and B. A. Roe.** 1980. Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. J. Mol. Biol. **143**:161–178.
36. **Sathymoorthy, V., and B. R. DasGupta.** 1985. Separation, purification, partial characterisation and comparison of the heavy and light chains of botulinum neurotoxin types A, B and E. J. Biol. Chem. **260**:10461–10466.
37. **Shone, C. C., P. Hambleton, and J. Melling.** 1985. Inactivation

of *Clostridium botulinum* type A neurotoxin by trypsin and purification of two tryptic fragments. Eur. J. Biochem. **151**:75–82.
38. **Simpson, L. L.** 1981. The origin, structure and pharmacological activities of botulinum toxin. Pharmacol. Rev. **33**:155–188.
39. **Simpson, L. L.** 1986. Molecular pharmacology of botulinum and tetanus toxin. Annu. Rev. Pharmacol. Toxicol. **26**:427–453.
40. **Simpson, L. L.** 1989. Botulnum neurotoxin and tetanus toxin. Academic Press, Inc., New York.
41. **Stechner, B., U. Weller, E. Habermann, M. Gratzl, and G. Ahnert-Higler.** 1989. The light chain but not the heavy chain of botulinum A toxin inhibits exocytosis from permeablised adrenal chromaffin cells. FEBS Microbiol. Lett. **255**:391–394.
42. **Sugiyama, H.** 1980. *Clostridium botulinum* neurotoxin. Microbiol. Rev. **44**:419–448.
43. **Tabor, S., and C. C. Richardson.** 1987. DNA sequence analysis with modified T7 DNA polymerase. Proc. Natl. Acad. Sci. USA **84**:4767–4771.
44. **Thompson, D. E., J. K. Brehm, J. D. Oultram, T.-J. Swinfield, C. C. Shone, T. Atkinson, J. Melling, and N. P. Minton.** 1990. The complete amino acid sequence of the *Clostridium botulinum* type A neurotoxin, deduced by nucleotide sequence analysis of the encoding gene. Eur. J. Biochem. **189**:73–81.
45. **Whelan, S. M., M. J. Elmore, N. J. Bodsworth, T. Atkinson, and N. P. Minton.** 1992. The complete amino acid sequence of the *Clostridium botulinum* type E neurotoxin, deduced by nucleotide sequence analysis of the encoding gene. Eur. J. Biochem. **204**:657–667.
46. **Yanisch-Perron, C., and J. Messing.** 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of M13mp18 and pUC19 vectors. Gene **19**:103–119.
47. **Young, M., N. P. Minton, and W. L. Staudenbauer.** 1989. Recent advances in the genetics of clostridia. FEMS Microbiol. Rev. **63**:301–326.